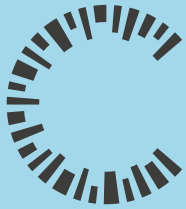


CISPA DISPLAY

EP 23

EN

INTRODUCTION

The pursuit of scientific excellence is in the DNA of the CISPA Helmholtz Center for Information Security. 39 CISPA-Faculty, our leading scientists, along with a constantly growing number of doctoral students, post-doctoral researchers and research group leaders are continuously striving toward this goal. Organized in six research areas, our researchers cover a wide range of topics including trustworthy information processing, cryptography, secure connected and mobile systems and trustworthy artificial intelligence. A look at the CSRankings, a metrics-based ranking of the best computer science institutions in the world, shows that their work has been crowned with success: CISPA regularly takes first place in the area of computer security, as it did in the past academic year.

Highly specialized knowledge within the research community

Scientific work is a highly specialized pursuit. The researchers at CISPA combine innovative application-oriented research with cutting-edge foundational research. Their results are shared and discussed primarily within the scientific community, especially at the major annual academic conferences, such as the USENIX Security Symposium, the ACM Conference on Computer and Communications Security (CCS) and the International Conference on Machine Learning (ICML). In order to speak at one of these conferences, researchers must submit substantial scientific papers in advance, which are then reviewed by a panel of experts. In computer science, accepted papers are then published at the conferences themselves rather than in scholarly journals, as is common in other scientific disciplines.

Sharing knowledge with the general public

To make CISPA research accessible to a wider audience including the general public, the Corporate Communications department at CISPA writes so-called ‘paper texts’ on a regular basis. These short, comprehensible summaries of scientific papers are mainly published on the CISPA website. They are based not only on the papers themselves, but also on personal interviews with the papers’ authors. Since scientific publications are often written by more than one researcher, these interviews are usually conducted with the paper’s first author or, in the case of

With the CISPA Display, we are adding a new print publication to the science-communication portfolio at CISPA. It brings together the ‘paper texts’ and the accompanying graphics that were created over the course of a year, intending to make some of the excellent research at CISPA accessible to the general public.

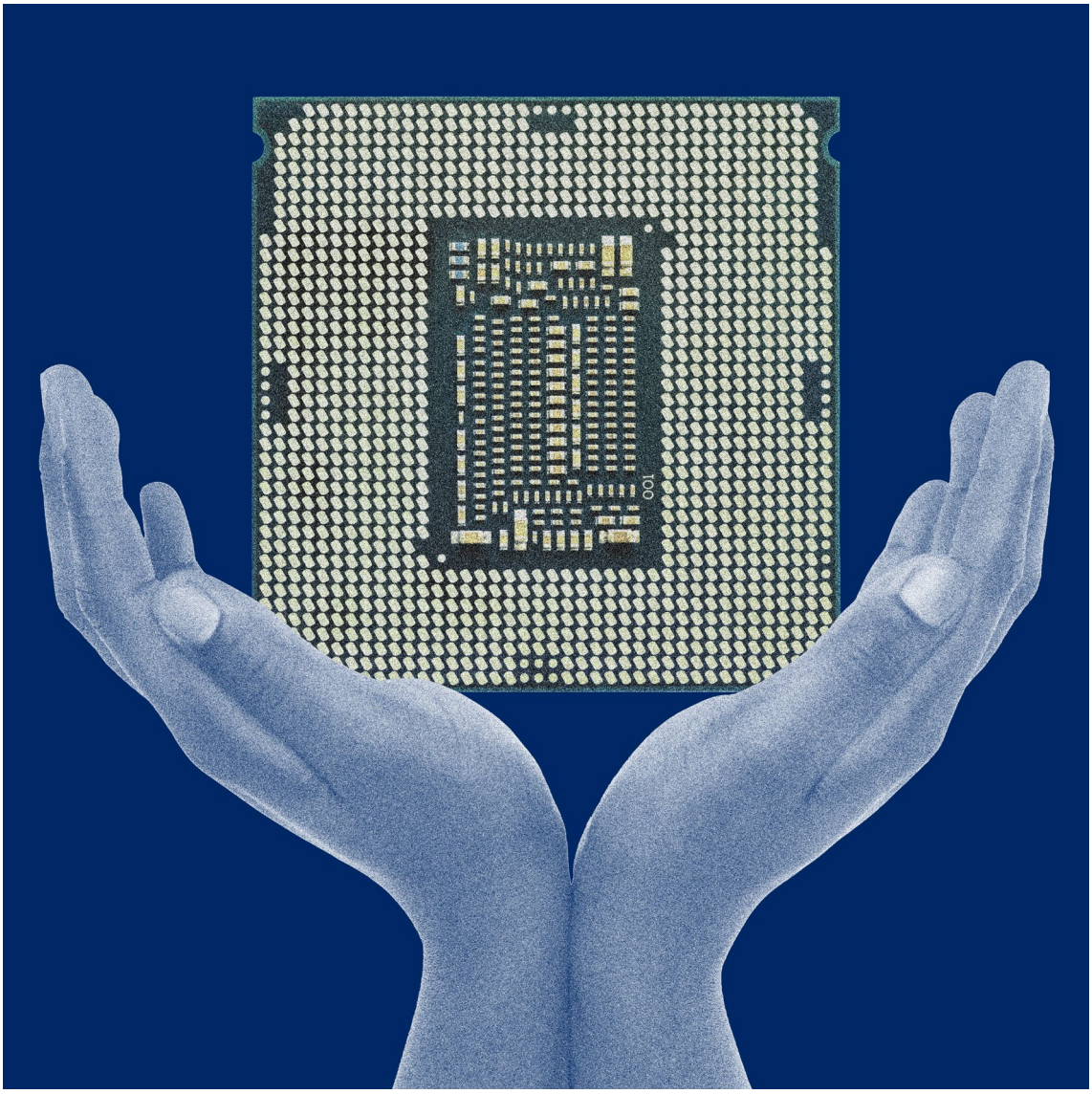
institutional collaborations, with the author affiliated with CISPA. For each of these texts, our communication designers create graphics that visualize the papers' key results or research questions, thus making a unique contribution to science communication at CISPA.

A new print publication for CISPA

Taken together, the paper texts and graphics are one of the cornerstones of science communication at CISPA. Their importance gave rise to the idea of dedicating a separate print publication to them: the CISPA Display. The CISPA Display 2023 brings together all the paper texts and corresponding graphics that were published in the previous year. Summarizing a total of 18 scientific papers by CISPA researchers, the publication now before you illustrates the great diversity of the research topics that are pursued at CISPA. They range from key management for cryptocurrencies, satellite security and authentication mechanisms in messenger services to methods for protecting against deepfakes. The CISPA Display expands our portfolio of regular print publications, which also includes the CISPA Zine. We hope you will enjoy reading it!

CONTENTS

CONTENTS



© Janine Wichmann-Paulus

In his paper "Automatic Detection of Speculative Execution Combinations", PhD student and CISPA researcher Xaver Fabian presents a new approach to automate the detection of vulnerabilities in processors – even if they only emerge through the combination of several speculative mechanisms. For his work, Fabian was honored with a Distinguished Paper Award at the Conference on Computer and Communications Security (CCS) in Copenhagen in November 2023.

New approach improves automatic detection of vulnerabilities in processors

Xaver Fabian

Processors are often described as the heart of a computer, but actually, they are the brain. The processor controls and interprets commands, coordinates processes, and ensures that tasks requested by users are forwarded to the appropriate places. Modern processors do this at enormous speed and perform many tasks in parallel. And there is a trick that helps them make good use of processor resources: speculative execution. Xaver Fabian explains what this means: “In phases in which the processor is not fully utilized, it uses various mechanisms to try to predict which program step could follow next. It then performs the necessary calculations and stores the results in a buffer. If the data is not needed after all, it discards it. However, it has been shown that the discarded data leaves a trace in the cache memory where attackers can read it under certain circumstances.” The first vulnerability to allow attacks of this type became known as ‘Spectre’ in 2018 and caused significant ripples. Researchers have been finding new Spectre variants regularly since then.

»One way of providing a security guarantee for this is to develop a formal model.«

Methods for closing Spectre vulnerabilities

After the first Spectre vulnerabilities became known, ad hoc measures were developed to close them. But who is to say these measures will actually work in all cases? “One way of providing a security guarantee for this is to develop a formal model. This makes a mathematical analysis possible and allows us to prove the effectiveness of the measures”, explains Fabian. Researchers create such models using logical mathematical methods and special formal languages. “So far, many researchers have done so by focusing only on certain variants of Spectre gaps and thus only on certain speculative mechanisms in the processor and looking at them in isolation. To be honest, this is because everything is complicated enough as it is. In reality, however, several speculation mechanisms are running in the processor simultaneously. We have tried to create a model that allows us to combine these mechanisms. Getting there was not easy. There are 200 pages of mathematical proofs alone somewhere on my desk.”

First, Fabian phrased two previously unexamined speculative mechanisms in formal language, thus making them accessible to mathematical proof. These formal expressions are called semantics. He then combined these semantics with an existing semantic for another speculative mechanism. “The basis for my work is provided by the preliminary work of IMDEA researcher Dr. Marco Guarnieri and others. Firstly, I used an assembly language they developed to formalize it. Secondly, I was able to incorporate the two new semantics I developed directly into the existing tool called SPECTECTOR for testing.” And his work paid off.

»The work we are doing here is absolutely foundational.«

Foundation for more complex analyses

Still, SPECTECTOR is insufficient for extensive testing of how well processors are protected against Spectre attacks. “The model is still quite simplified. Modern processors are very complex and can do quite a lot. The modeling approaches in research are still quite behind.” Nevertheless, his work lays the foundation for a far more complex analysis than has been possible to date. “The work we are doing here is absolutely foundational. However, that’s what’s needed. For a long time, too little attention was paid to IT security. But I have the impression that something has changed in recent years.” Fabian is delighted with his Distinguished Paper Award, which honors outstanding research papers. “I was honestly astonished. It’s great that my work is appreciated in the community.”

Fabian, Xaver; Guarnieri, Marco; Patrignani, Marco (2022) Automatic Detection of Speculative Execution Combinations. In: CCS 2022, 7-11 Nov 2022, Los Angeles, CA, USA. Conference: CCS ACM Conference on Computer and Communications Security

© Lea Mosbach

Visual digital certificates could be a secure and privacy-friendly approach to making official documents such as driver's licenses available digitally, says CISPA researcher Dañiel Gerhardt. "But only if they are properly verified for accuracy." Large-scale use of such certificates emerged during the pandemic in the form of Covid vaccination certificates, which were valid throughout the EU. In their paper, "Investigating Verification Behavior and Perceptions of Visual Digital Certificates", Gerhardt and his CISPA colleague Alexander Ponticello used this example to investigate why people often make mistakes when verifying digital certificates and how such mistakes can be avoided in future use cases. They presented their work at the prestigious USENIX Security Symposium 2023.

Why visual digital certificates are only secure in theory (so far)

Dañiel Gerhardt

Barcodes and QR codes have long been used to pass on information in a visually coded way. In everyday life, we encounter them, for example, on supermarket products, on packages, on concert tickets. “However, the amount of data that can be encoded in this form is limited. It is for this reason that, behind the QR or bar code, there is often only a link to an external source, such as a website”, Gerhardt explains. The Covid vaccination certificates, which were the ‘entry ticket’ to restaurants and other public places during the coronavirus pandemic, also featured QR codes. Behind these QR codes, however, was not just a link, but cryptographically signed data that could prove a person’s vaccination status. To do this, vaccination centers sent personal data such as a person’s name and date of birth, a personal identifier as well as the date of vaccination and information about the vaccine used to the Robert Koch Institute (RKI). The RKI in turn provided this information with a digital signature and issued a corresponding certificate. Vaccinated individuals could collect their certificates from pharmacies or doctors on presentation of their vaccination record and ID card. Certificates were available digitally and as paper copies. The QR code on the certificate could be scanned with apps such as the German ‘Corona-Warn-App’, which many people had installed on their cell phones. The RKI deleted any personal data once the signature was created. “The fact that the data was not stored centrally in Germany, but only locally, made the process very privacy-friendly. It’s also more secure against forgery, more sustainable and more cost-efficient, since no authority has to produce forgery-proof printouts”, Gerhardt explains.

The human factor affects verification

In practice, however, Gerhardt became aware of a problem: “When I went to a restaurant, for example, it happened a few times that instead of scanning the QR code and checking my ID, employees just took a look at the code in my app before letting me in. Obviously, that’s not a meaningful check. Others did use a scanning app suitable for verification and also scanned the QR code, but didn’t check my ID, for example.” Given that a viable

certificate verification process is critical to the certificates' security, Gerhardt conducted a qualitative interview study to explore why so many mistakes occur in practice. He observed and later interviewed 17 people who were responsible for verifying certificates as part of their jobs. "We wanted to get answers to two main questions: How do these individuals verify certificates and why do they do it this particular way? We also looked at how much the people carrying out the verification know about the verification process and how it works." Gerhardt's qualitative study is intended to facilitate a better understanding of user behavior so as to transfer the theoretical security benefits of visual digital certificates into the real world.

Gerhardt was surprised by the results: "The study participants checked the certificates in very different ways. I didn't expect so many variations." Some respondents performed all the necessary steps correctly during the verification process: They scanned the certificate with a suitable verification app, matched the data displayed in their app with the person's ID card, and also checked whether the photo really depicted the person standing in front of them. Other respondents also performed all of these steps, plus some unnecessary ones. "For example, some people tried to get a sense a of the person facing them and of their trustworthiness based on appearances. Others became instantly suspicious when presented with a screenshot, even though that's not really an indication that something dodgy is going on." Some respondents also showed distrust when they were presented with the certificate in an app they did not know. Other study participants tended to rely on their gut feeling rather than proper technical verification when assessing a vaccination certificate, and scanned the certificates only from time to time. Others stated that they only checked if a QR code was present and did not scan the certificates at all.

Better equipment and education required

"The majority of study participants didn't know much about the verification process and how it works technically. But that didn't necessarily mean they made mistakes in the process", Gerhardt says. "But it is also the case that those who understood the process well did not make any mistakes." In business, he says, other factors are often much more critical to how the verification process goes, for example, how time-consuming it is. "Also, some participants told us that their employer didn't provide them with a scanning device and they didn't want to use their personal smartphones for this purpose. Others, again, didn't know that they could simply download the verification app from the app store. They thought the app was only available to official bodies."

According to Gerhardt, these misunderstandingsoccured partly because many respondents did not receive

information from official bodies on how to check certificates. He thinks that better communication and education for those carrying out the verification process are required if the technology is to be used safely in the future. “Legislators could support that with a legal requirement for compliance.” It is also important to equip the verification personnel with appropriate test equipment and software, he said. “In addition, they need to know how to react when certificates don’t stand up to scrutiny.”

Clearer design, fewer misunderstandings

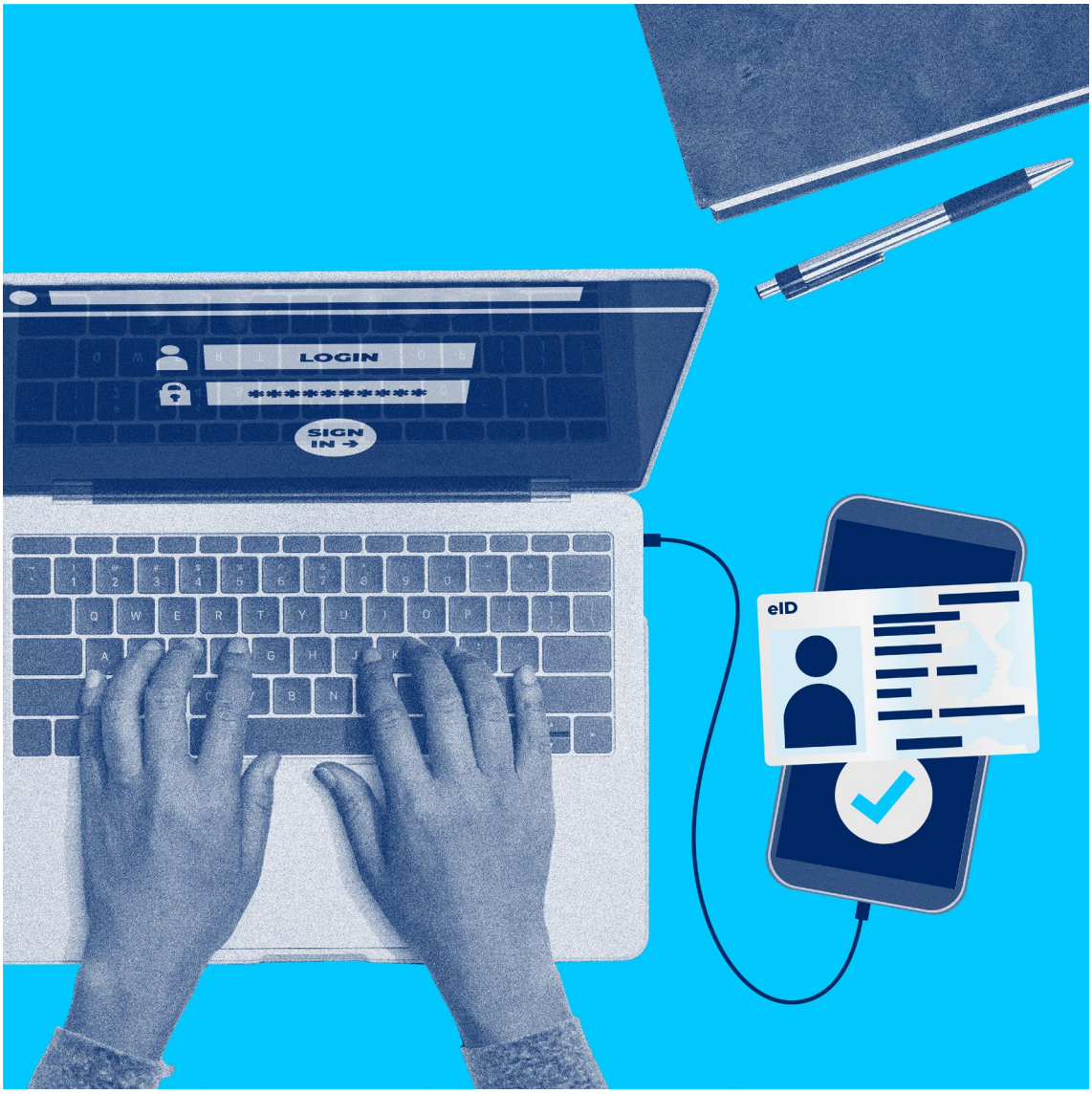
Last but not least, the design of the verification apps is an important consideration for Gerhardt. Some respondents were confused by certain details given in the app or even by the app’s color scheme. For example, some classified a certificate as secure as soon as the QR code was outlined in blue. Others were tempted to not carry out the verification process properly when the information “3 out of 3” (referring to the number of vaccinations received) was given in addition to the certificate. Gerhardt points out that care should be taken when designing future apps to prevent them from sending such unintended signals. He says: “If such measures improve the verification process and if visual certificates are implemented correctly, there are some useful areas of application for them. Digital driver’s licenses, for example, and electronic prescriptions. They could be digitally signed and securely issued by doctors.”

Off into the world of Usable Security

Gerhardt is thrilled that his paper was accepted at the prestigious USENIX Security Symposium. “The topic was already part of my Bachelor’s thesis. Together with CISPA researchers Alexander Ponticello, Adrian Dabrowski and Katharina Krombholz, I developed it into a full paper for the conference.” The 25-year-old is enrolled at the Graduate School of Computer Science at Saarland University. After the preparatory phase, he plans to do his doctoral research supervised by Krombholz. “I think the topics in the area of Usable Security are very exciting and I really enjoyed working together with the researchers at CISPA.”

Gerhardt, Dañiel; Ponticello, Alexander; Dabrowski, Adrian; Krombholz, Katharina (2023) Investigating Verification Behavior and Perceptions of Visual Digital Certificates. In: 32nd USENIX Security Symposium, 9-11 Aug 2023, Anaheim, CA, USA. Conference: USENIX Security Symposium

Researcher: Dañiel Gerhardt
Author: Annabelle Theobald



© Janine Wichmann-Paulus

2-factor authentication has become the standard for logging in to many web services. While many users use a combination of password and cell phone codes, the FIDO2 standard is considered to be the most secure variant but requires an additional hardware component. CISPA researcher Fabian Schwarz and his colleagues from the teams of CISPA-Faculty Professor Dr. Christian Rossow and CISPA-Faculty Dr. Lucjan Hanzlik have now developed FeIDo, a new method that does not require special user hardware. They presented their paper "FeIDo: Recoverable FIDO2 Tokens Using Electronic IDs" at the renowned ACM Conference on Computer and Communications Security (CCS) in 2022.

Developing an open-source prototype for 2-factor authentication

Fabian Schwarz

It's a simple fact: Without a log-in, many areas of the World Wide Web and in particular a large number of services, from messengers and information services to online banking, are not available to users. With every new account, however, users give away data and have to remember new passwords. At the same time, it is widely known that passwords are a rather insecure way of logging in, which is why a large number of new methods is being used and tested. This is where Fabian Schwarz's considerations came in. "We wanted to make the login process in web services as simple as possible and at the same time as secure as possible", he explains. The goal of Schwarz and his colleagues was to make previous standards both available and securely usable to the general public. Their focus was on the FIDO2 standard for 2-factor authentication, which was developed by the international FIDO Alliance. FIDO is the abbreviation for "Fast Identity Online".

The special feature of the FIDO2 standard is that it relies on additional hardware components for authentication. This can be a security token in the form of a USB stick, which may additionally be protected with a fingerprint scanner, or a smartphone with the latest security standards. FIDO2 uses the W3C Web Authentication Standard (WebAuthn) and the Client-to-Authenticate Protocol (CTAP) of the FIDO Alliance. Authentication is performed with both a private and a public key, which are generated by the security token. While the private key never leaves the security token, the public key is stored on the servers of the respective web services being used. Users enter the private key to request authentication, which can be securely verified and matched by the web services using the public key.

Disadvantages of previous methods

The FIDO2 standard allows passwords to be supplemented by the use of hardware-based security tokens. In the long term, FIDO tokens such as the YubiKey by the Yubico company aim to enable completely passwordless authentication. Although Schwarz thinks that this is a welcome development, he also believes that there are also a number of drawbacks. There is, for example, the cost factor. Users have to acquire new hardware components, for example in the form of hardware security

tokens or smartphones with the latest security standards. In addition, there is a negative impact on user-friendliness that comes with the high security standard. If the security token, such as a USB stick for example, with the login data is lost, it is no longer possible to log in and access to the user's online accounts is blocked. Procedures for restoring access exist, but they usually have security gaps, Schwarz explains, or involve additional user setup, such as the upfront registration of a backup token. The challenge for Schwarz in developing a new method was to reduce these drawbacks.

From FIDO2 to FeIDo

The starting point for Schwarz and his colleagues is a simple but all the more compelling idea: using things that almost everyone has at their disposal, such as an ID card and a cell phone. "We looked at how electronic ID cards or passports might be used for this without sensitive user data contained in the passports going to the website operators", Schwarz explains. They wanted to take advantage of the fact that modern cell phones can also read eIDs via NFC technology, i.e. contactless data transmission using radio waves. All that is required is an NFC-enabled smartphone, which includes almost all commercially available Apple and Android cell phones, but no extra hardware. "All that is then needed is a small intermediate app that carries out the reading process and transmits data to our specially protected service", Schwarz continues. This is exactly what the researchers implemented as a prototype. They then successfully subjected it to various theoretical security tests.

Anonymous log-in as an expanded field of application

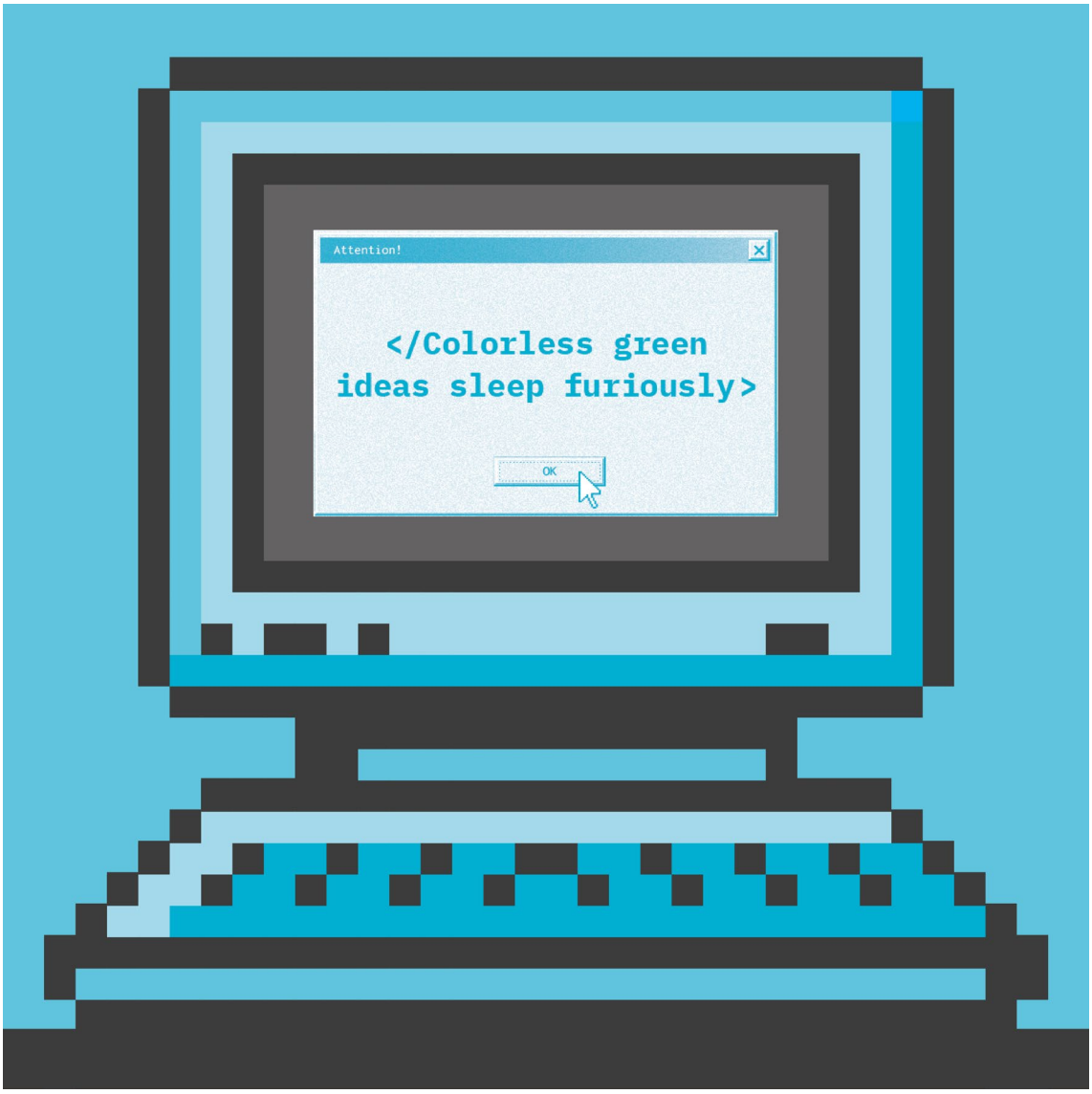
Schwarz and colleagues see even more advantages in the FeIDo process which result from working with data from eIDs. Decisively, in the FeIDo process, this data is read out but not passed on. This distinguishes FeIDo from other processes that also use personal data from eIDs for authentication. This makes new fields of application for FeIDo conceivable, such as checking age restrictions when logging into specially protected web services. "We can use our app to enable anonymous login, but at the same time our service provides proof that the user is of age", Schwarz explains. In order to use this variant, however, changes would have to be made to the applications of the web services. "However, this would not cause any problems," Schwarz continues. For web services without add-ons such as age verification, the log-in procedure of the CISPA researchers' prototype could be used immediately.

Good feedback at CCS conference

Schwarz and his colleagues first presented their paper at the ACM Conference on Computer and Communications Security (CCS), which was held in Los Angeles in

November 2022. The topic was met with a great deal of interest: “There were so many questions that there wasn’t enough time to answer them all”, Schwarz recounts. In response, Schwarz and his colleagues also published an extended paper. Schwarz himself has since turned to other research topics. However, the results of his research, specifically the prototype application, are open source and freely available. “We designed the whole thing to be openly usable as a community project, like a Tor browser”, he explains. The goal is for the community to provide such a service free of charge. Schwarz would be delighted to see colleagues or companies take on the project and develop the prototype further.

Schwarz, Fabian; Do, Khue; Heide, Gunnar Hanzlik, Lucjan; Rossow, Christian (2023) FeIDo: Recoverable FIDO2 Tokens Using Electronic IDs (Extended Version). *Technical Report. UN-SPECIFIED.*



© Lea Mosbach

A new specification language called ISLa is set to enhance automated software testing or fuzzing. ISLa might also become a milestone in the field of software security and reliability. In a paper entitled "Input Variants", CISPA researcher Dr. Dominic Steinhöfel has described his recent work. He also presented it at the renowned European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE) in November 2022. ISLa is an important component of CISPA-Faculty Professor Dr. Andreas Zeller's research project "S3 - Semantics of Software Systems". Together with his team, Zeller aims at making any software in the world amenable to automated testing. The ERC funds Zeller's research project with an ERC Grant of 2.5 million Euros.

New specification language is a game changer for automated software testing

Dominic Steinhöfel

In the programming of software, errors are frequent occurrences. To prevent these from causing system crashes or security gaps, researchers often test their programs prior to release with the help of so-called fuzzers. “These tools produce huge numbers of random inputs to see how a program will perform in live operation. However, it is difficult to produce inputs that are capable of testing the deeper programming functions”, Steinhöfel explains.

The basic structures of the data languages spoken by computers and used to phrase programming input resemble those of human languages. Which is why not only grammar, but also semantics plays an important role in all of this. Using a striking example, US-American linguistic Noam Chomsky illustrated the difference between the two in the 1950s: “Colorless green ideas sleep furiously.” While this sentence is syntactically impeccable, it is semantically incorrect. The grammatical structure may be perfectly fine, but still the sentence makes no sense at all.

The thing about semantics

When a fuzzer produces an input that is grammatically correct but that contains no meaningful message for the program being tested, the input is rejected by the parser. The parser is a sub-program that checks if the input is intelligible for the program. If this is the case, the parser converts it into a format that is adequate for processing. If the input is unintelligible, however, the parser produces an error message and disregards it entirely. “With inputs like this, you can only test the quality of the parser but not the stability of the program itself”, Steinhöfel explains. There are fuzzers that produce smarter inputs and thus circumnavigate the parser. “This is where the process often ends because this is also where the more complicated characteristics at the semantic level come in.”

New specification language is key

ISLa, Steinhöfel’s new specification language, can become a game changer in this context. “ISLa allows us to understand inputs with a precision that was previously unknown and hence to test programs deeply

and thoroughly." According to Steinhöfel, the key lies in a very general formalism that makes almost any program accessible. "But we do need an input description. We can write it manually or else learn it from an existing program." This, however, is complicated and oftentimes only possible in an approximate manner. "There will always be programs that are too large or too complicated to be understood completely. But we can continue to become better at it."

For this, ISLa is a powerful tool: Not only can it generate inputs, it can also test, repair and mutate them. What is more, ISLa allows researchers to describe a program's output. "If we can describe the input as well as the output, we can describe the behavior of the entire program. This allows us to do an awful lot: We can determine how a program is meant to behave, we can analyze how it does behave and we can force it to behave it in the way we want it. In short: If you control the input as well as the output, you control the program." CISPA Faculty Andreas Zeller, with whom Steinhöfel has closely collaborated, emphasizes the importance of Steinhöfel's research: "ISLa opens up entire worlds for the testing of systems."

»ISLa allows us to understand inputs with a precision that was previously unknown and hence to test programs deeply and thoroughly.«

From theory to practice

In the near future, Steinhöfel will draw on the basis provided by ISLa to develop practical approaches for the testing of relevant software systems. Among other things, he will focus on the learning of complex input and output descriptions and concentrate on state-based systems such as databases and servers. Further, he intends to ascertain whether it is possible to combine already established testing methodologies.

A PostDoc researcher in the research group of CISPA-Faculty Professor Dr. Andreas Zeller, Dominic Steinhöfel earned his degree as well as his PhD from TU Darmstadt. “Without knowing exactly where it would lead me, I have been working towards ISLa since 2021.” In these words, the pride in his achievement is shining through. And rightly so.

Steinhöfel, Dominic; Zeller, Andreas (2022) Input Invariants. In: Technical Track, 2022. Conference: ESEC/FSE European Software Engineering Conference and the ACM SIGSOFT Symposium on the Foundations of Software Engineering (formerly listed as ESEC)

© Lea Mosbach

Humanitarian aid programs are deployed in difficult, sometimes even hostile environments where there is usually no adequate digital infrastructure. Also, aid recipients often have little agency protecting their interests in the process of obtaining humanitarian aid. This urgently calls for the development of scalable solutions for aid distributions that do no harm. In collaboration with the International Committee of the Red Cross, CISPA-Faculty Dr. Wouter Lueks and his colleagues from EPFL in Lausanne have developed a new privacy-friendly solution for large-scale humanitarian aid distribution. His paper has recently been published at the prestigious 44rd IEEE Symposium on Security and Privacy, where it was honored with a Distinguished Paper Award.

A new a token-based system for humanitarian aid distribution combines accountability and privacy

Wouter Lueks

In 2021, 3,575,484 people worldwide received food assistance from the International Committee of the Red Cross (ICRC). Aid organizations strive to assist victims of violence, famine and disaster in regions with limited internet connectivity. And they do so using limited financial resources. To ensure that aid organizations can help as many people as possible, the distribution process must be efficient and accountable. Traditionally, humanitarian organizations use different forms of paper-based systems to support aid-distribution. These, however, are difficult to scale to large groups of recipients. They also complicate audits when it comes to verifying that donor money was well spent. To address these challenges of efficiency and auditability, organizations have recently started exploring digital solutions. Most of them integrate so-called Identity Management systems (IdM), as commonly used in passports. However, the use of IdM-based solutions brings significant privacy risks to the vulnerable population of aid recipients: Personal information stored in central databases might leak or be abused.

Requirements for a solution

The ICRC approached Dr. Wouter Lueks about a privacy-friendly solution for these problems. “I was triggered for two reasons”, he explains: “It’s a technical challenge and it’s a privacy-sensitive topic.” The project that evolved is based on a collaboration between Lueks’ former employer, the EPFL in Lausanne and the ICRC. “The usual approach would have been to use biometrics, because fingerprints don’t just change”, Lueks continues. For the ICRC, however, using biometrics was not the preferred option. Biometric data are extremely privacy-sensitive precisely because they do not change. Also, securing these data is difficult. At this point, Lueks’ research interests came into play. “Typically, we build a system to solve a problem. I’m interested in disentangling the risks that materialize when you design systems to solve that problem. Some risks are inherent in the problem itself,

others come from how you design the solution." One possible risk of organizing fair distribution on the basis of fingerprints, for example, is that the underlying central database could be used by state and non-state actors to identify groups of recipients and subject them to repression. Here the risks stem from a design choice. In order to mitigate such risks, Lueks and the ICRC worked closely together to design a better solution. Two workshops and regular meetings were held over the course of one year. The outcome was a list of requirements for possible solutions that ranged from deployment conditions to security and privacy factors, while ensuring that the ICRC's ethical standards were fulfilled.

A token-based aid-distribution system

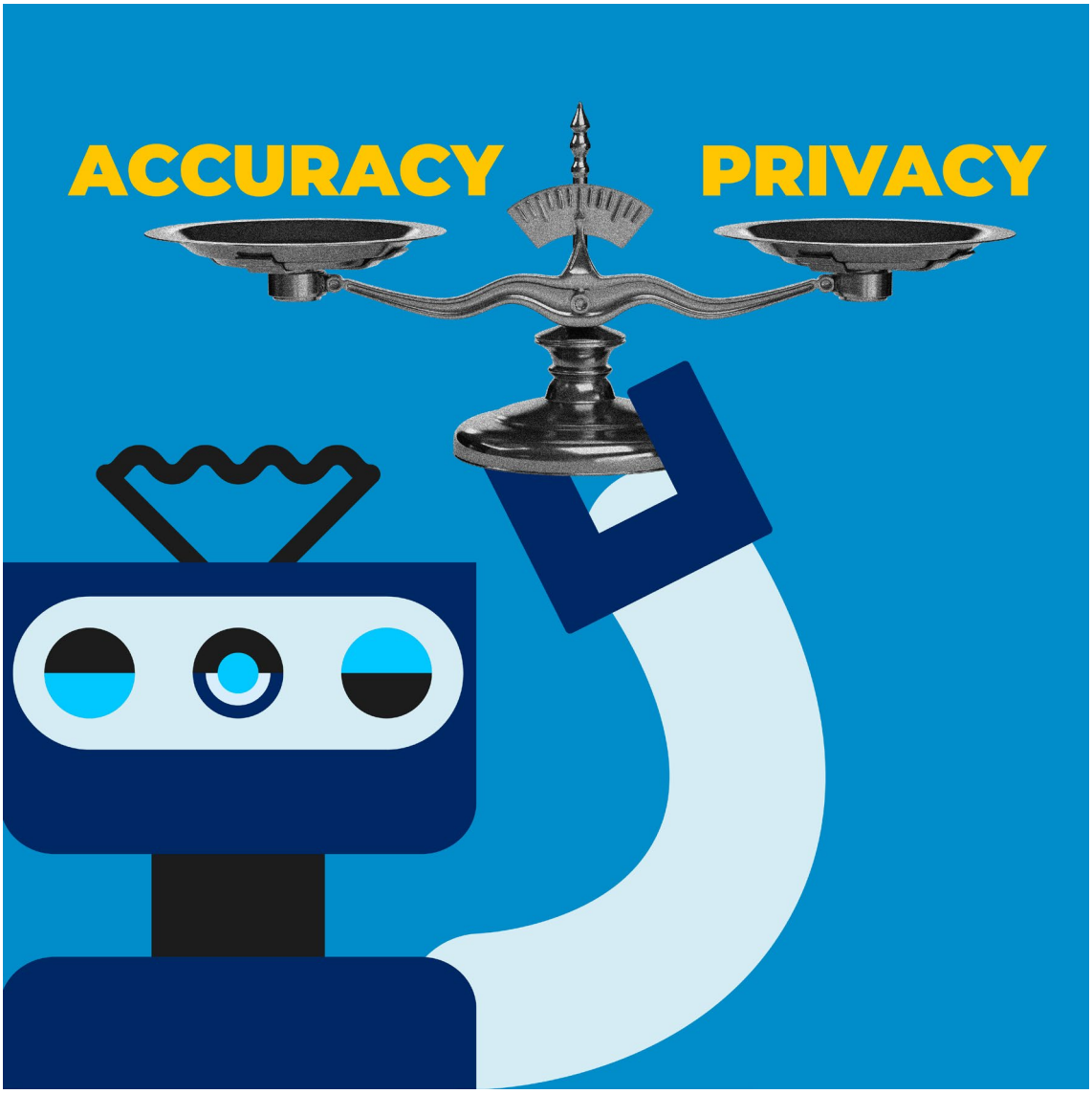
Lueks and his fellow researchers came up with a token-based approach to satisfy the aforementioned requirements. The most important design choice was to decentralize information using digital tokens, meaning that all collected information is stored only on a token that stays with the recipient. The token can either be a smart card or a smartphone. Smart cards have the advantage of being cheap and suitable for large-scale operations where digital infrastructure is lacking, while phones are easier to deploy (if available). The token-based scheme follows the existing humanitarian aid distribution workflow (see figure below). Once the system is set up, which can be done outside the target region and before the start of a mission, there is no need for updates or internet connectivity. The tokens work offline, meaning that the smart cards communicate locally with registration and distribution stations upon presentation. "One of the key challenges in this design", explains Lueks, "was how to ensure that only eligible persons can receive aid and that audit records cannot be faked, while at the same time revealing as little information as possible about aid recipients." Using Lueks' token-based design, the distribution station and auditor can verify the eligibility of any recipient, while no information about the recipient themselves is being revealed. The design thus ensures privacy, while maintaining auditability.

More efficiency increases capability to help

Implementing a decentralized aid-distribution system that relies on digital processes can help increase the efficiency of humanitarian aid operations. More efficient registration and distribution processes can save time and money on the ground and thus increase the NGOs' capabilities to help people in need. Until now, the problem with paper-based solutions, as well as many digital solutions, was a lack of privacy. The solution created by Lueks closes this gap. "In the process of figuring out what the real problem is you often uncover new challenges, which makes this work very satisfying." Lueks explains.

His approach of focusing on the problem rather than the solution, fits the Do-No-Harm approach implemented in humanitarian aid. This approach aims at identifying unintended negative as well as positive impacts of humanitarian interventions before the start of a mission. Lueks has demonstrated that this also applies to the design of new digital solutions. In the future, he would like to continue to work with NGOs: “For me, they are interesting partners because they work to benefit society.” And using technology for a good cause is what Lueks is striving for.

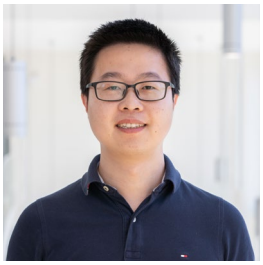
Wang, Boya; Lueks, Wouter; Sukaitis, Justinas; Graf Narbel, Vincent; Troncoso, Carmela (2023) Not Yet Another Digital ID: Privacy-Preserving Humanitarian Aid Distribution. In: 44th IEEE Symposium on Security and Privacy, May 22-25, 2023, San Francisco, CA, USA. Conference: SP IEEE Symposium on Security and Privacy



© Lea Mosbach

Differential Privacy is considered a gamechanger for privacy protection in data analysis. The technique has been in use for a number of years. The US Census Bureau, for example, used Differential Privacy (DP) to conduct the last census data publication in 2020. The tech giant Apple, which is particularly committed to data protection, also relies on DP to analyze the data of its users in a privacy-compliant manner. CISPA research-group leader Zhikun Zhang explains what DP is, what problems it still faces, and how he wants to use his research to further improve the process.

The new gold standard: Rethinking Differential Privacy



Zhikun Zhang

It's a well-known fact that data is a tradable commodity in the digitized world. Not everyone, however, is aware of how well data collection and analysis already serve society and how much more they could contribute in the future. A few examples: Experts believe that the analysis of medical data such as blood values, oxygen saturation, MRI scans or X-ray images with the help of artificial intelligence (AI) will take healthcare to an entirely new level over the coming years. AI can combine and analyze huge amounts of data. Autonomous driving also would be inconceivable without the processing of immense amounts of sensory data that are collected all around and inside the car. Not to mention such common conveniences as forecasts for when the public swimming pool will be least crowded or where the next traffic jam is likely to be. All of this is only possible through the analysis of huge amounts of data.

A lot of data, a lot of protection

These examples also make it easy to see where the problem might lie: Much of the data mentioned is enormously sensitive as it reveals quite a bit about us, our state of health, our habits and movement patterns. The protection of privacy, already a well-known issue in itself, is thus becoming more relevant today than ever before. Since 2006, a solution appears to have been found with Differential Privacy. "The new gold standard of privacy protection is Differential Privacy", says Zhikun Zhang. According to Zhang, the goal of Differential Privacy (DP) is actually quite simple: to learn as much as possible about a specific group of people from an existing dataset, without learning anything about the individuals in that group.

What's behind differential privacy?

"For one thing, the term provides a mathematical definition of privacy. It's a kind of statistical guarantee that individual people's data won't affect the outcome of queries on larger datasets", Zhang explains. "On the other hand, it's also often used to describe the specific process by which database queries are answered in a way that maintains privacy." DP was first developed by the cryptographer Cynthia Dwork. Together with fellow researchers, she introduced the first formula for measuring how much of a privacy violation a person faces when their data becomes part of a larger data collection and, thus, public.

Noise for more privacy

The large amounts of data collected today are mostly used to train machine learning models to perform a variety of tasks. For example, a model based on a large dataset from cancer patients, comprising for instance blood values, genetic information and MRI findings, could be trained to detect developing cancer much earlier than is currently the case. To ensure that highly sensitive medical data such as this remains secure, it must be anonymized in some form. It is not enough, however, to remove personally identifiers such as names or addresses. Multiple queries and the combination of characteristics that at first glance appear to be of little significance often allow unambiguous conclusions to be drawn about individuals. Instead, so-called 'noise' is introduced to the data. This involves a number of methods that are used to produce a kind of controlled randomness in the response to queries.

Still many challenges for research

The important thing is for data processing to still retain its statistical utility despite this noise. And that's not the only challenge. Often, a number of special algorithms need to be employed, and accesses need to be logged and kept on record, because too many queries could reveal too much, even with noisy data. Artificially produced data with strong privacy guarantees may offer the solution to these problems. "We publish such synthetic data that meet DP standards and reflect the statistical properties of the real datasets, but that are not subject to the same limitations in processing."

According to Zhang, the challenge in creating synthetic data under DP is to identify the most informative statistical information possible. That's the only way to extract as much useful data as possible, even from complex datasets, such as those that map people's movement patterns or their social connections within networks. He has published several papers on his research, including presentations at the prestigious USENIX Security Symposium.

Multifaceted topic

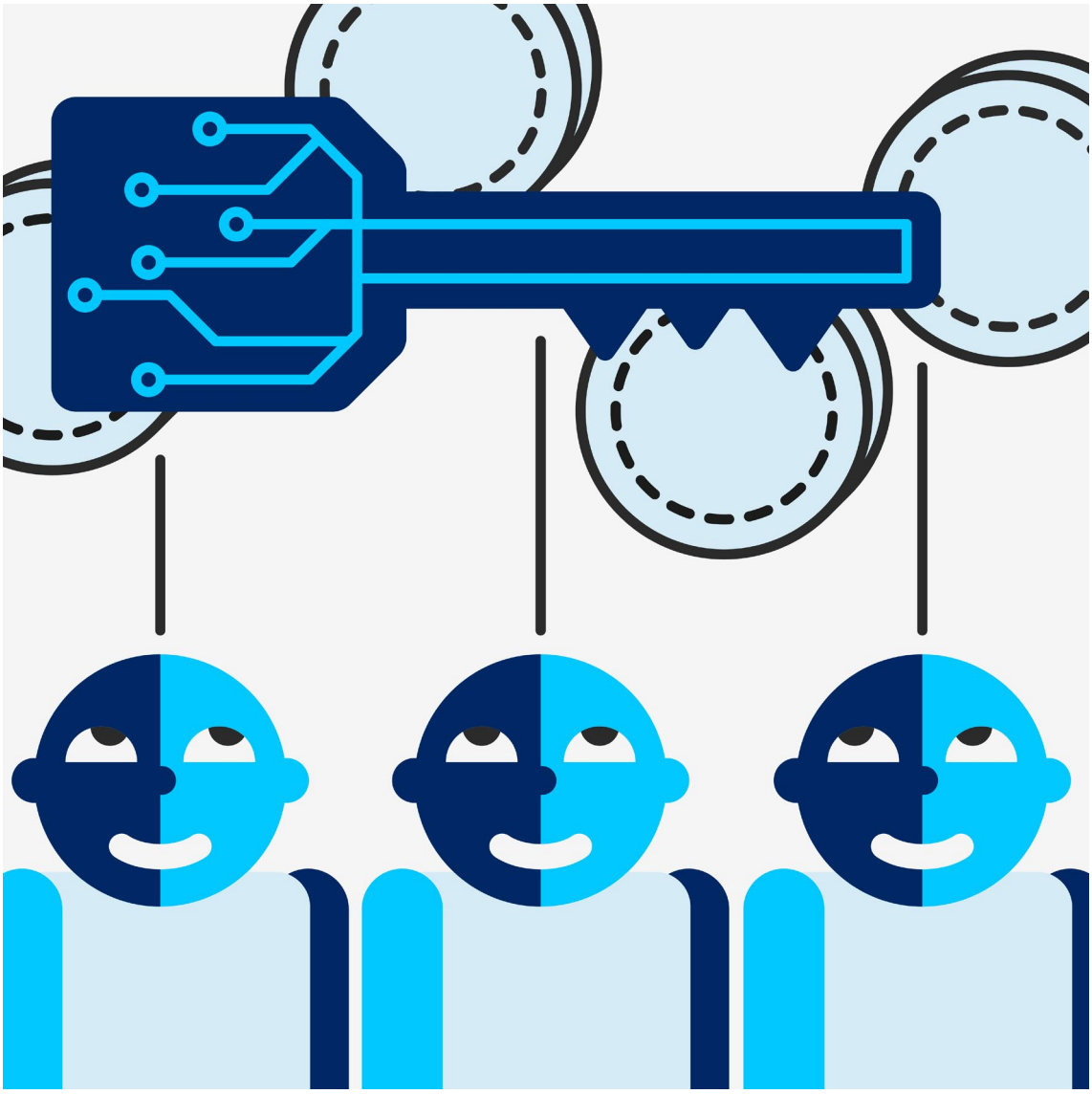
Zhang has been conducting his research beneath the California sun since October 2022. "I am a participant in the CISPA-Stanford program and am currently a visiting professor at the University of Stanford." Differential privacy continues to be a topic that keeps him busy. "I'm currently doing research with a colleague at Stanford on the question of privacy protection within large-language models, such as those in Chat-GPT, and what impact the use of differential privacy might have on such models." It feels like he might strike gold.

»We publish such synthetic data that meet DP standards and reflect the statistical properties of the real datasets, but that are not subject to the same limitations in processing.«

Wang, Haiming; Zhang, Zhikun; Wang, Tianhao; He, Shibo; Backes, Michael; Chen, Jiming; Zhang, Yang (2023) PrivTrace: Differentially Private Trajectory Synthesis by Adaptive Markov Model. In: 32nd USENIX Security Symposium, 9-11 Aug 2023, Anaheim, CA, USA. Conference: USENIX Security Symposium

Zhang, Zhikun and Wang, Tianhao and Honorio, Jean and Li, Ninghui and Backes, Michael and He, Shibo and Chen, Jiming and Zhang, Yang (2021) PrivSyn: Differentially Private Data Synthesis. In: 30th USENIX Security Symposium, 11-13 Aug 2021, Vancouver, B.C., Canada. Conference: USENIX Security Symposium

Researcher: ***Zhikun Zhang***
Author: ***Annabelle Theobald***



© Janine Wichmann-Paulus

Until a few years ago, only very tech-savvy and venturesome speculators invested in Bitcoin and the like, but cryptocurrencies are slowly becoming a new asset class on conventional financial markets. To trade cryptocurrencies securely, owners require cryptographic keys, which they need to keep secret. Key management systems have been developed for single users. In financial institutions, however, more than one person need to have access to these keys. CISPA researcher and PhD student Carolyn Guthoff conducted a qualitative survey with 13 financial professionals. She shows how key management systems need to be adapted for new applications in the financial sector. Guthoff presented her paper entitled "Perceptions of Distributed Ledger Technology Key Management" at the prestigious IEEE Symposium on Security and Privacy (S&P).

Key management is a challenge for crypto funds

Carolyn Guthoff

A decentralized currency to which no bank, state, or authority has access – that is the idea behind Bitcoin. First described in a 2008 document, Bitcoin is still the best-known use case of the so-called distributed ledger technology (DLT), but it is by no means the only one. “Distributed ledger” means exactly what is says: The term refers to a database for transactions that is stored on many computers. Thus, it is not managed from one single place, but rather by many users in a decentralized fashion. “There has been an absolute hype around distributed ledger technologies in various industries since 2014”, Guthoff says.

Far more familiar to many people than DLT is the term “blockchain”. Blockchain is one of the best-known distributed ledger technologies. It provides also the basis for cryptocurrencies. “The name comes from the fact that, in a blockchain, blocks of data are stored one after another. Blockchain applications such as Bitcoin or Ethereum are based on the same technology, but they follow different rules”, Guthoff explains. The goal, however, is always to have a currency that facilitates online payments entirely without the involvement of any financial institution.

»The results of this study can help design key management solutions that meet the needs of financial institutions.«

Single-user scenario for cryptocurrencies is outdated

With this original idea in mind, it is hardly surprising that a service and management system has emerged around cryptocurrencies which is geared toward individual users. This is also true for the management of cryptographic keys, which is essential for the processing of transactions in a blockchain. Every financial transaction between two trading partners on the blockchain has to be documented in detail. It is also visible and traceable for all users. This is the only way to ensure that the system as a whole remains trustworthy and reliable. In addition to a public key, cryptocoin owners also own a private key with which they can access their digital wallet and digitally sign transactions. These private keys are 52 characters long and randomly assigned to users. If such a key is lost, the cryptocurrency associated with the key is also irretrievably lost. Secure storage of private keys is therefore essential.

Study provides design ideas for key management in multi-user scenarios

The requirements for secure management and storage of cryptographic keys grow when multiple users need to have access to the keys. “By default, this is the case for crypto funds, for example. Since a change in the law in 2022, such funds have become a bigger issue in financial institutions in Germany and, as with other funds, they are usually managed by several employees. In addition, employees must be able to cover for each other in the event of vacation or illness”, Guthoff explains. She conducted a survey with 13 employees in financial institutions enquiring about the security and confidentiality requirements the institutions had for key management as well as the optimal key management envisioned by employees. “The results of this study can help design key management solutions that meet the needs of financial institutions, that are secure yet practical.”

»Respondents overwhelmingly wanted technical solutions for storing keys that could be secured by multiple factors.«

In practice, employee turnover is one of the biggest challenges for key management with multiple users. “Once an employee had access to a key, there is the risk that they have copied it. In that case, they can still access the assets, even if they have resigned in the meantime or hold a different position”, Guthoff says. A good solution to this problem, according to some respondents, might be a program that allows keys to be used for transactions but that does not allow direct access to the key itself. “Across the board, respondents overwhelmingly wanted technical solutions for storing keys that could be secured by multiple factors, such as TANS and passwords.”

Another important question is how to deal with liability and responsibility issues. Many of the respondents think that models for optimal key management and for the allocation of access rights to assets should correspond to the organizational structure of their company. “This means, for example, that CEOs have access to higher assets than ordinary employees and should be given extended access rights”, Guthoff explains. Most respondents also wanted key management that did not require too much background knowledge of digital signatures and that was easy to use. Some respondents found it useful to involve an intermediary between the financial institution and the trading platform who would be responsible for key management and its security, provided there was an appropriate relationship of trust.

Many exciting research questions

For CISPA researcher and PhD student Carolyn Guthoff, this was the first paper she submitted to a conference. “That my paper was accepted at the prestigious S&P is wonderful. It encourages me.” Despite her recent deep dive into cryptocurrency, Guthoff is not planning to focus her research on financial topics in the future. “Working on this topic was super exciting, but now I will turn to other research questions. I’m particularly interested in topics where the ideas and demands of security researchers and the realities of users’ lives don’t really fit together.” There are probably still a few of those.

Guthoff, Carolyn; Anell, Simon; Hainzinger, Johann; Dabrowski, Adrian; Krombholz, Katharina (2023) Perceptions of Distributed Ledger Technology Key Management - An Interview Study with Finance Professionals. In: 44th IEEE Symposium on Security and Privacy, 22-25 May 2023 San Francisco, CA, USA. Conference: SP IEEE Symposium on Security and Privacy

Researcher: Carolyn Guthoff
Author: Annabelle Theobald

In 2022, more than 1.14 billion websites were online worldwide. Many of these are either hosted in the European Union (EU) or frequented by people from the EU. Since 2018, the European General Data Protection Regulation (GDPR) applies to these websites. It obliges businesses and website operators to ensure the protection of their customers' and users' personal data. Together with her colleagues, CISPA researcher Christine Utz has investigated how website operators can be alerted to an inadequate implementation of the GDPR and ePrivacy Directive on their sites. They have published their findings in a paper entitled "Comparing Large-Scale Privacy and Security Notifications".

Website operators take security more seriously than data protection

Christine Utz

When Internet users visit a website via a web browser, there usually occurs an exchange of data. For example, website operators often track the IP address from which their website is accessed. Personal data is also often provided by users themselves, for example when they purchase products online and have them delivered to their homes. The General Data Protection Regulation (GDPR), which has been in force since 2018, is the first EU-wide standardized guideline for processing personal data. It aims at protecting users from excessive data storage. Storing an IP address, for example, is already considered storage of personal data. The GDPR applies to all websites hosted or accessible in the EU. Website operators are responsible for implementing the directive, while the national data protection authorities are responsible for its supervision.

In 2019, CISPA researcher Christine Utz and her colleague Martin Degeling from Ruhr University Bochum investigated how websites had changed after the introduction of the GDPR. “Our main finding was that while there had been little change in the actual practice of tracking, there had been an increase in transparency efforts by websites, for example, via the provision of privacy statements as well as the introduction of cookie banners”, Utz recounts. This was one of the starting points for their current study. CISPA Faculty Dr. Ben Stock, in whose group Utz is a researcher, had conducted an earlier study on how e-mail campaigns could be used to inform website operators about security vulnerabilities. “This led to the idea of investigating whether website operators could also be made aware of a lack of data protection with the help of such a campaign”, Utz continues.

Study design and approach

After extensive preliminary research, the actual implementation of the study took place with a set of approximately 160,000 websites. The criterion for including a website in the sample was the existence of a data protection problem such as the absence of a data protection statement, the absence or delayed display of a cookie banner, or existence of input fields for personal data without HTTPS protection. As a comparison criterion, unsecured access to a so-called Git repository was included in the study. A Git repository is a working copy of a website

stored on an external server. Early in November 2021, the website operators were automatically contacted by e-mail and informed about the problems. Over a period of two months, Utz then observed whether the problems on the sites were fixed or not, both among those who had been contacted as well as among a control group. In order to gain deeper insights into why actions were taken or not, the researchers also attached a questionnaire to their e-mails and examined the e-mail communications with the website operators.

Challenges during implementation

A study with such a large sample entails a number of challenges, some of which result from the automation of work steps. One risk concerns false positives. Automated tools used to search HTML source texts, for example, may fail to detect existing privacy statements due to inconsistent naming. Another hurdle is the selection of e-mail addresses. Previous studies have shown that the use of generic addresses such as 'info@-' or 'webmaster@-' has disadvantages. For this reason, wherever possible, e-mails were sent to specific e-mail addresses recognized on the website. "The biggest difficulty, however, was to prevent our e-mails from being classified as spam by the recipients", Utz explains. To do so, Utz and her colleagues took a number of precautions. They used an external server for hosting and the e-mails were also signed. The external server was also intended to prevent all e-mails coming from CISPA from being classified as spam, which could have caused reputational damage to the center.

Pitfalls of large-scale notification campaigns

The most important finding of the study was that, in principle, it is possible to use large-scale e-mail campaigns to alert website operators to data privacy problems. Nevertheless, given the immense resources required to conduct such a study, success in terms of problem resolution is quite limited. This is particularly evident from the fact that only a very small proportion of those informed responded to the e-mails. The percentage of websites on which changes were made during the observation period was in a low single-digit range. A comparative analysis also showed that security vulnerabilities are more likely to be remedied than data protection problems. One reason for this, Utz believes, is the fact that security vulnerabilities can often be addressed with less effort.

The qualitative examination of the questionnaires and e-mail communications revealed further reasons for the limited campaign success within the observation period. Utz found that website operators were less open to notifications about privacy issues than about security breaches. Further obstacles to the implementation of changes were also identified. These included language barriers due to a lack of English language skills on the

part of the e-mail recipients as well as the classification of the notification e-mails as spam. Interestingly, the GDPR reference itself also proved to be an obstacle. Some operators doubted whether their own website even fell within the scope of the GDPR, or the reference to a lack of data protection was simply rejected as inapplicable. The participants would have liked more, and more detailed, information on data protection issues.

The goal is to cooperate with data protection authorities

Utz wants her research to increase the enforceability of the GDPR. “Data protection authorities often don’t have the capacity to detect inadequate GDPR implementations on websites and to point these out to website operators”, she says. “But we as researchers could support the authorities in doing this.” Importantly, researchers and the institutions behind them have the necessary technical and human resources for such projects. Conversely, researchers could benefit from the standing of public authorities. “Data protection authorities can communicate more effectively why the GDPR is important”, Utz says. Cooperation would therefore be a win-win situation for all involved. However, according to Utz, it is essential that broad-based notification campaigns via e-mail are accompanied by other measures in the future, such as information campaigns about the scope of the GDPR. She also suggests the implementation of a new standard concerning the availability of website operators so that e-mail notification campaigns lead to greater success. This could be done, for example, by means of a privacy.txt file stored on all websites, which contains information on how the operators can be contacted in the event of data protection-related questions.

Utz, Christine; Michels, Matthias; Degeling, Martin; Marnau, Ninja; Stock, Ben (2023) Comparing Large-Scale Privacy and Security Notifications. In: PETS 2023, July 10–15, 2023, Lausanne, Switzerland. Conference: PETS Privacy Enhancing Technologies Symposium

Researcher: Christine Utz
Author: Felix Koltermann

While spaceflight to the moon and farther beyond has always attracted much public attention, the real conquest of space is taking place, silently, in Low Earth Orbit (LEO). At a distance of between 200 km-1,000 km, LEO is rather close to the earth and contains a rapidly increasing number of relatively small, relatively cheap satellites. In their paper "Space Odyssey: An Experimental Software Security Analysis of Satellites", CISPA Faculty Ali Abbasi, Thorsten Holz and their research team investigate the security issues that accompany the dawning of this "New Space Era". At the IEEE Symposium on Security and Privacy in May, their publication was awarded a Distinguished Paper Award, an honor given only to the top 1 percent of submitted papers.

Space oddities: Examining satellite security

Ali Abbasi

It is a hallmark of the New Space Era that the number of satellites orbiting the Earth is on the rise. A great proportion of these is made up of LEO satellites, whose small size and cost make them accessible not only to nation states and large corporations but also to small institutions and businesses. Amazon, for instance, provides satellite communications on-demand, renting out ground stations as a service. Orbiting Now, a website gathering satellite information, counted 7,004 active LEO satellites in mid-May 2023. Depending on their payload, these satellites can perform different missions, among which are Earth observation, weather forecasting, navigation, communications as well as space science.

It was this sudden, wide accessibility of LEO satellites that sparked Ali Abbasi and Thorsten Holz's research interest. "There is a paradigm shift happening. And whenever there's a paradigm shift, there are security issues", Abbasi says. The conviction, long-held by satellite engineers, that obscurity granted security is no longer valid, as he explains: "For a long time, the assumption was that satellites weren't accessible and that, therefore, they were secure. But LEO satellites have lots of connectivity features." The lack of official security standards for satellites is an additional worry, as Holz adduces: "You can only contact a satellite via a proprietary radio protocol. But the frequencies on which they are communicating are not regulated."

Space oddities: Examining satellite security

Satellites are controlled by, and communicate via, a communication mechanism called 'bus'. It comprises the Communications Module (COM), which receives radio messages from the ground station, and the Command and Data Handling System (CDHS), which processes and executes any incoming commands. If the COM can be seen as the satellite's ears, the CDHS functions as its brain: It carries a computer platform which operates on the basis of preinstalled, onboard software. Satellites are thus similar to other, more common computer systems and similarly vulnerable to software attacks. Hypothesizing that satellite systems would be less secure than modern Windows, Linux or MacOS systems, the researchers focused their efforts on the attack surfaces provided by satellite firmware.

As a starting point for their examination, they drew up a taxonomy of possible threats against satellite firmware, identifying three overarching attacker goals and sketching all possible attack paths that might be used for their realization. From an attacker's point of view, the ultimate goal may be to compromise the availability of the satellite, to gain access to satellite data or else to seize control of the entire satellite. This last attacker goal has at the same time the greatest potential for damage: If a satellite is seized and used to attack another, the debris resulting from the crash may cause a domino effect in which space becomes cluttered with loose satellite parts. Called the Kessler Syndrome, this effect is in Abbasi's words largely "Hollywood stuff".

"Once you have access, it's just too bad": Uncovering firmware vulnerabilities

Hollywood stuff or not, the research team led by PhD student Johannes Willbold successfully triggered error conditions in the CDHS, seizing full control of two out of three real-life satellites in the applied part of their study. Their case studies were carried out on three real-life, in-orbit LEO satellites. After liaising with the institutional owners of these satellites over a protracted period of time, they acquired the satellites' firmware images for the purposes of a security analysis.

The results yielded by these case studies underline the fact that research on satellite security has been a long time coming. The most important of their findings concerned the security of the COM. As the entry point for radio messages from the ground station, the COM should ideally function as a gatekeeper, keeping out suspicious commands. If it fails to fulfil this role, the CDHS can be assailed by unforeseen input. If this input then succeeds in triggering an error condition in the onboard software, it effectively interferes with the satellite's brain.

»There's a paradigm shift happening. And whenever there's a paradigm shift, there are security issues.«

Even though satellites are highly complex systems, the software vulnerabilities uncovered by the researchers are surprisingly standard. As Holz points out, “In the Linux or Windows world, we have studied software faults of this kind for many years. But in these embedded systems, the defenses are 20 years behind of what we know from commodity systems.” As it is, the most effective defense mechanism actually lies outside the system, as Willbold highlights: “The barrier here is access. But once you have access, it’s just too bad.”

Responsible disclosure: Promoting satellite security

The researchers reported all of the software issues they detected to the owners of the three satellites well before they published their study. This procedure, called responsible disclosure, is part and parcel of their professional code of conduct and it is also indispensable for the promotion of satellite security. Going forward, these systems can only be protected if researchers, operators and developers begin to cooperate, as Abbasi highlights: “Those who shared their satellite firmware with us are really brave. They really care about cybersecurity. In the short term, they have gained nothing. There may be a problem with their software, but all software has problems. But in the long term they helped protect space systems.”

Willbold, Johannes; Schloegel, Moritz; Vögele, Manuel; Gerhardt, Maximilian; Holz, Thorsten; Abbasi, Ali (2023) Space Odyssey: An Experimental Software Security Analysis of Satellites. In: 44th IEEE Symposium on Security and Privacy, 22-25 May 2023 San Francisco, CA, USA. Conference: SP IEEE Symposium on Security and Privacy

Researcher: Ali Abbasi
Author: Eva Michely

CISPA-Faculty Dr. Michael Schwarz has been researching side-channel attacks for years. He was involved, among other things, in the discovery of Platypus and Meltdown. These are cyberattacks in which data is stolen via a detour, the so-called side-channel. Side-channel attacks exploit information that the Central Processing Unit (CPU) reveals involuntarily during processing, such as run-time behavior or power consumption. With Collide+Power, Michael Schwarz, his PhD student Lukas Gerlach, and a group of researchers at TU Graz have now discovered a new power side-channel attack that directly targets the CPU and that can theoretically hit all processors.

Collide+Power: New side-channel attack affects all CPUs

Michael Schwarz

With Collide+Power, attackers can extract data directly from the processor. This is because all data that is processed by a computer system has to pass through the Central Processing Unit (CPU), which contains short-term memories or caches. Here, data that has already been processed is stored temporarily so that it can quickly be retrieved and reused. When data stored in the cache is overwritten by new data, for example because users access another password in their password manager, power is consumed. At this point, a physical effect comes in: The more data is changed in the cache, the more power is required.

Data collide in the cache

Collide+Power takes advantage of this effect. The malicious code programmed for the attack fills the cache with data known to the attacker. If the user now accesses a program – such as their password manager – the attacker's data in the cache is overwritten with the password: Attacker and user data 'collide' in the cache. The power consumption of the CPU during the overwriting process allows the attackers to draw conclusions about the password. "The more similar the attacker's data and the data from the target program are, the less power is consumed – and power consumption can be measured very accurately", explains Schwarz.

»The more similar the attacker's data and the data from the target program are, the less power is consumed.«

Of course, many different computing processes take place in parallel in the caches of a computer, for example because various programs are open at the same time. So how can an attacker identify the part of the calculations in the cache that they want to exploit? “The injected malicious code reloads the data from the program under attack countless times in the cache”, Gerlach points out. These constantly repeated loading processes allow the attacker to draw conclusions about the data records that are relevant to them.

Power consumption allows conclusions to be drawn about data

This type of data theft is possible because, in computer memories, all values are represented based on a binary code. Each individual value is coded with multiple digits, each of these being either a 1 or a 0. For one byte, which has eight digits, the number 1 would be represented by “0000 0001”, the number 2 by “0000 0010”. Thus, to overwrite a 1 in the cache with a 2, two digits, namely the last two, have to be changed. If a 1 is overwritten with a zero, which is represented by “0000 0000”, only the last digit is changed. This requires less power. By comparing the amount of power consumed with each change, Collide+Power manages to ‘guess’ each of the individual digits of a value.

Many repetitions of this ‘guessing process’ are necessary to capture every digit of a value and, thus, the secret. This makes the process very complex and time-consuming. With the current malicious code, extracting a credit card number, for example, would take 4 – 5 hours, the researchers estimate. “However, this is only our test code. If you are serious about this, you could surely optimize the code”, Schwarz says.

Collide+Power closes a research gap

Collide+Power closes a gap in the detection of power side-channel attacks. It is the first side-channel attack that uses power measurements to derive data directly from the processor. Since the hardware itself is targeted by Collide+Power, it is impossible to prevent this kind of attack. Manufacturers can only provide information and notifications. So far, Michael Schwarz says, Collide+Power has not been seen in practice: “As researchers, we can only show that the attack is possible. How dangerous it is, is for the manufacturers to judge.” However, Lukas Gerlach adds, “you lose the guarantee that data will remain untouchable.”

»As researchers, we can only show that the attack is possible. How dangerous it is, is for the manufacturers to judge.«

Kogler, Andreas; Juffinger, Jonas; Giner, Lukas; Gerlach, Lukas; Schwarzl, Martin; Schwarz, Michael; Gruss, Daniel; Mangard, Stefan (2023) Collide+Power: Leaking Inaccessible Data with Software-based Power Side Channels. In: 32nd USENIX Security Symposium, 9-11 Aug 2023, Anaheim, CA, USA. Conference: USENIX Security Symposium

© Janine Wichmann-Paulus

“As a cellular-network researcher, you are in some ways a prisoner of your geographic position”, says CISPA researcher Dr. Adrian Dabrowski. What he is referring to is that, because of the large number of providers and networks, cellular-network researchers have so far only been able to carry out tests and measurements in foreign cellular networks at immense expense. Together with Gabriel Gegenhuber from the University of Vienna and other colleagues from SBA Research, Dabrowski has therefore developed MobileAtlas, a kind of infrastructure that allows testing from any location across Europe. He presented his paper “Geographically Decoupled Measurements in Cellular Networks for Security and Privacy Research” at the renowned USENIX Security Symposium 2023.

MobileAtlas: Mapping mobile communications security

Adrian Dabrowski

2G, 3G, 4G, 5G – what sounds a little like a bingo draw actually refers to the mobile communication standards currently in use. The latest 5th generation standard – for 'generation' is what all the Gs stand for – is still being developed. The oldest of the bunch, 2G, was introduced back in the 1990s and is still in use. "2G is mainly used for voice transmission or for simple smart devices; such as a beverage dispenser that indicates it needs to be refilled", Dabrowski explains. The subsequent 3G was switched off in Germany in 2021 and replaced by 4G, also known as LTE. With 4G, users can enjoy streaming services or make video calls on the go, for example. These mobile communications standards, which exist side by side, apply worldwide. Roaming is meant to enable mobile-network customers to use the services agreed with their mobile-network provider when abroad while at the same time delivering the promised security and privacy protection. At least that's what it's meant to to.

Is the 'Roam-Like-At-Home' principle an empty promise?

At issue here is the so-called 'Roam-Like-At-Home' principle promised to EU citizens in the EU Roaming Regulation, which was revised in 2022. The Bundesnetzagentur (the German Federal Network Agency) writes: "As a result of the revision of the Roaming Regulation, not only will the same price apply when traveling in the EU as at home, but also basically the same quality." Dabrowski doubts that this promise can be kept: "With roaming, the home network and the network of the country I'm

»If you look closely, there is no consistency between roaming and non-roaming connections.«

visiting work together. They want to offer a service that is also as consistent in terms of privacy and security as the one in the home network. But the technical implementation is completely different." For example, when on vacation in Switzerland, the voice connection is established directly via the Swiss network, while the Internet connection takes a detour via Germany. In the home network, both would go the direct route. "If you look closely, there is no consistency between roaming and non-roaming connections", Dabrowski explains. Mobile providers have an extremely large amount of leeway, he says, and have been virtually impossible to control. According to Dabrowski, this also applies to network security.

Cross-border tests hardly possible so far

The problem: So far, tests and measurements across borders have been extremely costly. Dabrowski elaborates: "Europe is extremely fragmented. There are many mobile network providers in each country. Germany, with only three providers, is the exception. If I find that there is a security gap in one of our domestic mobile networks and want to check whether this is also the case in other networks, I currently have two options: Either I travel around a lot and test every network in every country in every constellation, or I equip as many devices as possible in every country with as many different SIM cards from different providers as possible. In no time at all, I thus end up with 1,000 SIM cards, 1,000 contracts and a private bankruptcy."

"Decoupled measurements" are the solution

The solution might come in the shape of a framework that the researchers have developed and that allows the geographical separation of the SIM card from the cellular modem. The modem is a component in mobile devices such as smartphones that establishes the connection between the device and a cellular network. Its job is to put the radio data into the right form and send it to, and receive it from, transmission towers. The SIM card serves to identify the user and to assign the smartphone to a specific network. Dabrowski explains the connection between all this and his framework: "Normally, the SIM card and the phone are one unit. We split up this unit and remove the SIM card from the phone. We simulate the communication protocol over the Internet and are thus able to travel virtually. Let's take an example to simplify this: We connect the SIM card to our measuring station in Germany and pretend to be in Germany. Then we disconnect it and connect it to our measuring station in France and pretend to be there. All we need for that is a device in Germany or else in France."

Cost-effective and open source

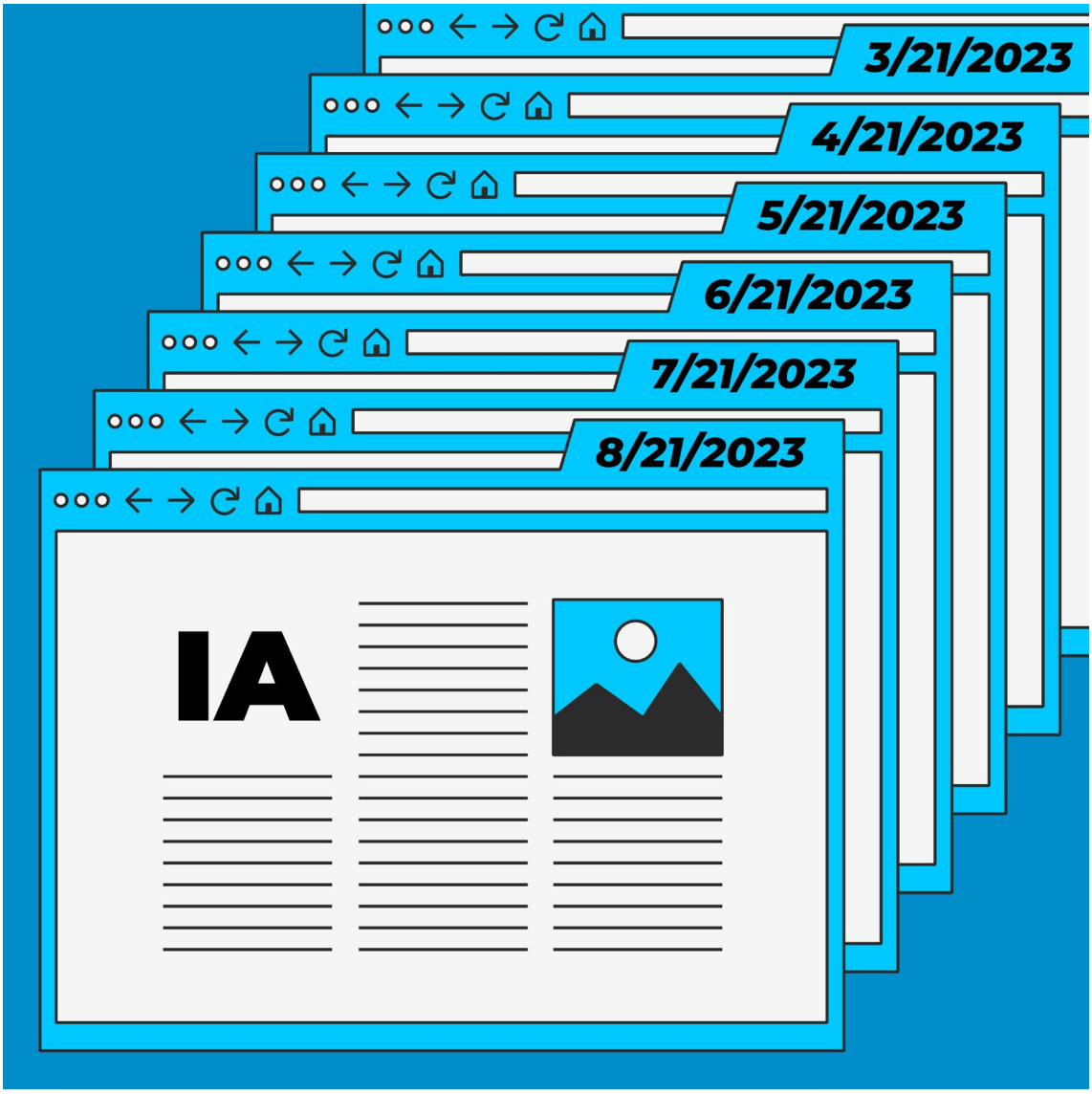
According to Dabrowski, their measurement and test platform, which works for standards 2G to 4G, provides a

controlled experimentation environment that is extensible and cost-effective. “In addition, our approach is open source, so other researchers can contribute sites, SIM cards and measurement scripts.” The researchers are making the platform accessible and usable under the name MobileAtlas. The tool is likely to be of interest not only to researchers. “Mobile operators could also use it for the first time to check whether their roaming partners keep their promises.” The name MobileAtlas is no coincidence. According to Dabrowski, it was derived from the name of the RIPEATLAS Internet test platform, which has been in existence since 2010. “RIPE NCC is the European Internet Governance. The RIPE Atlas is a global network of meters that measure the connectivity and accessibility of the Internet.”

The boundaries of borderlessness

With MobileAtlas, measurement stations have been set up in ten countries so far, along with the suitable infrastructure for the measurements. Dabrowski hopes that the measurement network will quickly expand with the help of other researchers. “However, we will also have to make sure that no mischief is done with the SIM cards so that we don’t incur any costs. It remains to be seen Whether we can offer MobileAtlas as comprehensively as RIPE NCC offers its platform.” Dabrowski and his colleagues have already shown that their approach can uncover interesting information: “We have discovered, for example, that certain services can be camouflaged in some mobile networks in such a way that the data traffic they generate is not deducted from the data volume included in the rate. Worse for end users, however, are the security problems that we were also able to demonstrate. In some cases, we found problematic firewall configurations or uncovered hidden SIM card communication with the home network.” The findings are not all that alarming in this regard. Exploiting these problems would require very targeted attacks and savvy attackers. “But gaps like that are never good. And now we have the opportunity to point them out to vendors.”

Gegenhuber, Gabriel Karl; Mayer, Wilfried; Weippl, Edgar; Dabrowski, Adrian (2023) MobileAtlas: Geographically Decoupled Measurements in Cellular Networks for Security and Privacy Research. In: 32nd USENIX Security Symposium, 9-11 Aug 2023, Anaheim, CA, USA. Conference: USENIX Security Symposium



© Lea Mosbach

Quality in science is measured, among other things, by whether studies can be reproduced by other scientists and still yield the same result. This, however, is a major challenge, especially for studies on Internet security mechanisms. After all, websites, which provide the underlying data, are constantly changing entities. CISPA researcher Florian Hantke and his colleagues from the team of CISPA-Faculty Dr. Ben Stock along with researchers from Ca' Foscari University in Venice have taken a new approach to tackle this problem: Using web archives instead of live analyses for these studies, they have achieved promising results.

A new standard? Using web archives for live analyses of website security

Florian Hantke

Studies on website security have an important place in the research area of information security. To this day, the research standard is live analysis. This means that website security parameters are measured the moment that researchers access a website. The problem is that this always represents only a snapshot: What is 'live' one moment may be out of date a moment later. "The web is so random that it is extremely complex to reproduce experiments", says CISPA researcher Florian Hantke. For this reason, it is almost impossible to repeat experiments under the same conditions in live analyses.

For Hantke, this poses a fundamental problem: "Experiments should always be reproducible, because otherwise they lose relevance. Otherwise, anyone could simply claim that the Internet is safe." According to Hantke, one alternative that could theoretically guarantee reproducibility is the use of web archives. At regular intervals, web archives store copies of existing websites, so-called 'snapshots', on external servers. There, they can be retrieved along with dates and time codes. Unlike live websites, the stored copies are not subjected to changes anymore. The best-known web archive is the Internet Archive. In research, web archives have so far been used mainly for historical analyses, not for live analyses. Hantke explains this by saying that "many people think that archives do not contain all the important data."

Internet Archive superior to other web archives

Hantke and his colleagues wanted to know how well web archives were suited for live analyses of website security mechanisms. To ascertain this, they had to find out which of the existing web archives stored the most accurate copies. They examined a set of public web archives in terms of volume and quality of the deposited data for the 5,000 most important websites in the period from January 2016 to July 2022. In a comparison of these web archives, the Internet Archive (IA) showed the best results. The quality of the archive is so good that, under certain circumstances, Hantke and his co-authors even recommend working with IA as the sole source.

The researchers verified the data quality of IA in a case study of two mechanisms that are standard on many websites: security headers and Java Script inclusions. They were also able to show that IA stores copies of websites with such regularity that it allows for even more detailed analyses, the quality of which equals the quality of live analyses. In addition, IA allows for the analysis of multiple snapshots of a website over the same time period, which Hantke calls “neighborhood”. This allows any short-term outliers in the data, such as a website’s server problems, to be smoothed out. The approach used by the researchers of using publicly available web archives makes it easier for studies to be reproduced. In the long term, this can increase the research quality and make it easier to check security mechanisms of websites.

The challenges of using web archives

Nevertheless, there are also some things to consider when using web archives for live analytics. “One major disadvantage is the slow speed”, Hantke explains. For example, processing large amounts of data is much faster in a classic live analysis because access to data stored in web archives is very slow. However, researchers could solve this problem by establishing collaborations with the archives they favor in order to get better access to the data. “The different vantage points also need to be considered”, Hantke continues. These are the access points from which websites are accessed around the world. These access locations determine what exactly a website looks like when it is stored in the archive. “For security issues, the differences tend to be negligible, but for analyses of the implementation of the DSGVO, for example, the access location is important”, he explains. This is because specific features relevant under the General Data Protection Regulation (GDPR) are often only displayed on European websites. In this case, a copy stored in the U.S. would not be of help. This is why, for each new research question, it has to be ascertained whether working with web archives is an option.

Productive PhD research

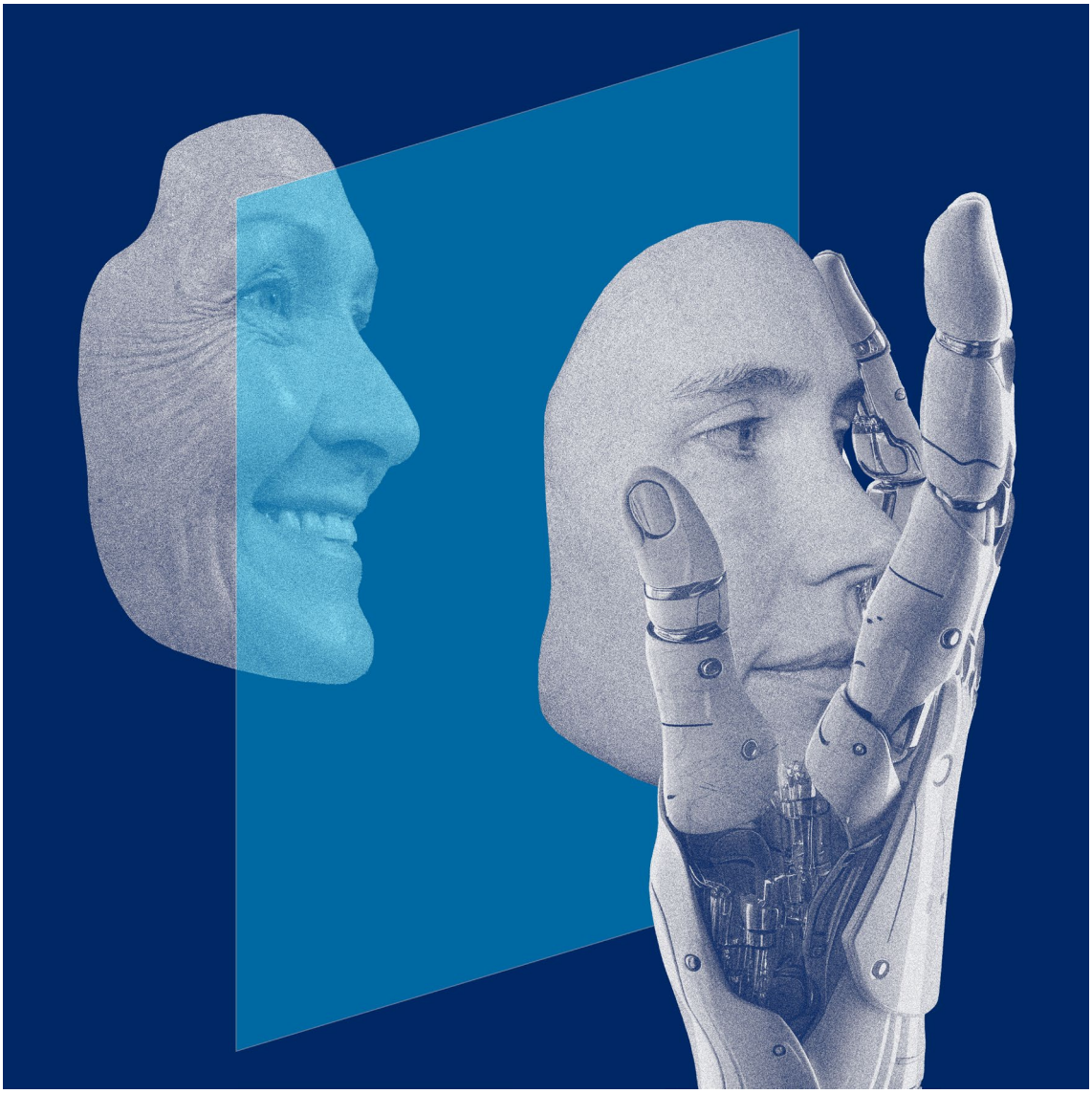
Florian Hantke is a PhD student and has been working at CISPA for a year now. He lives in Erlangen with his wife and works from home a lot. Asked whether he needs any special research equipment at home, he explains that a secure VPN connection to the CISPA server in Saarbrücken is entirely sufficient. “I can simply send an instruction to the server and run the analyses there”, Hantke says. He can then retrieve the results at a later point. The paper on web archives is already his second publication. “I’m quite happy with my output”, he admits with a laugh. For the summer, he is already planning another paper. But before that, he hopes there will be more interest in his findings on using web archives for security analyses. In

any case, the management of Internet Archive has already signaled interest. Together with his co-authors from Ca' Foscari University in Venice, Hantke is also planning a publicly accessible project for web security analyses, which other researchers will also be able to use.

»The web is so random that it is extremely complex to reproduce experiments.«

Hantke, Florian; Calzavara, Stefano; Wilhelm, Moritz; Rabitti, Alvise; Stock, Ben (2023) You Call This Archaeology? Evaluating Web Archives for Reproducible Web Security Measurements. In: ACM CCS 2023, 26-30 Nov 2023, Copenhagen, Denmark. Conference: CCS ACM Conference on Computer and Communications Security

Researcher: ***Florian Hantke***
Author: ***Felix Koltermann***



© Lea Mosbach

The omnipresence of images on the internet on the one hand and the exponential learning curve of AI image generators on the other heighten the risk of image manipulations with malicious intent. CISPA researcher Zheng Li and his colleagues have tested a technique that can partially prevent this. The results of their research have been published in a paper called "UnGANable: Defending Against GAN-based Face Manipulation" at the renowned "USENIX Security '23" conference.

Testing a new technique to safeguard against deepfakes

Zheng Li

A significant feature of contemporary online communication is the exchange of images on social networks. Once uploaded, these images often remain online for a very long time, which paves the way for manipulation and misuse. In addition to identity theft, where real images are used to create fake accounts, images might also be submitted to AI image generators and used for deepfakes. Deepfakes are image manipulations that cannot be detected by the human eye. This is a real risk especially for politicians and celebrities. “Most of the time, the people in these images are not even aware of the manipulation and cannot do anything against it”, CISPA researcher Zheng Li explains. This opens the door to disinformation: “This is why deepfakes pose a real threat to democracy.” Deepfakes can be generated by a number of AI-based mechanisms, for example by so-called GANs.

GAN is short for Generative Adversarial Network and refers to a machine learning model. GANs consist of two artificial neuronal networks that communicate with one another. Simply speaking, one of these two networks generates new data, images for example, while the other

»Most of the time, the people in these images are not even aware of the manipulation and cannot do anything against it.«

compares this new data to an existing data set, assessing the differences between the two. This assessment is fed back to the generating network, which uses it for improvements, for example to increase the similarity between the generated image and the real image. This also improves the algorithm itself. As the image resolution becomes better and the look ever more photorealistic, more possibilities emerge to manipulate images of real people. This primarily concerns 'face manipulation' which tampers with individual characteristics such as facial expressions or hair color. An important method for the generation of deepfakes is GAN inversion. It is a special technique for image processing in AI image generators.

Developing UnGANable

"Our starting point was the realization that up to now it has been impossible to prevent deepfakes that are based on GAN inversions", Li says. "That is why we decided to call our technique UnGANable", he recalls. "Simply speaking, UnGANable tries to protect images of faces against deepfakes." GANs can only process images if they first convert them into mathematical vectors or so-called 'latent code'. This is called GAN inversion and amounts to a kind of image compression. Using the latent code of a real image, the generator is able to create new images that are deceptively similar to their real-life precursor.

The procedure developed by Li and his colleagues obstructs GAN inversions and thus hinders attempts at forgery. At the level of mathematical vectors, UnGANable produces maximum deviations called 'noise' that are invisible and that hinder the conversion into latent code.

»Our starting point was the realization that up to now it has been impossible to prevent deepfakes that are based on GAN inversion.«

The GAN effectively runs dry because it cannot find any data that it might use to create new images: If no copies of the original image can be created based on the latent code, then image manipulation is simply not possible. Test runs with UnGANable with different GAN inversion techniques yielded satisfactory results. Li and his colleagues were also able to prove that their method offers better protection than alternative techniques such as the program Fawkes. Fawkes, developed by a research group at the Sand Lab in Chicago, is based on a distortion algorithm which conducts changes at pixel level that are invisible to the human eye.

Areas of application

Li's work is an important step toward new defense systems against face manipulation. "Protecting people against the malicious manipulations of their images matters to me", Li declares. The code of UnGANable is open source and available to other researchers. Those well-versed in handling code can already use it to protect their images against misuse. The general public, however, will have to wait until the corresponding software is programmed. Li has already turned to other projects but some of his colleagues continue their research on related topics. They want to ascertain, for example, if the technique behind UnGANable can also be used for other AI-based procedures that generate images from textual input. "Perhaps the technique can also be applied to videos in the future", Li hopes. In any case, the exponential learning curve of GAN-based techniques for image generation and manipulation will make the development of defense systems ever more necessary.

Li, Zheng; Yu, Ning; Salem, Ahmed; Backes, Michael; Fritz, Mario; Zhang, Yang (2023) UnGANable: Defending Against GAN-based Face Manipulation. In: 32nd USENIX Security Symposium, 9-11 Aug 2023, Anaheim, CA, USA. Conference: USENIX Security Symposium

Researcher: Zheng Li
Author: Felix Koltermann

© Janine Wichmann-Paulus

Messenger services offer a relatively high level of security through standard end-to-end encryption. However, that's only true as long as the right person is actually doing the chatting on the other end. Few people realize that the authentication of their chat partners is crucial to prevent attacks on the messaging process. In a self-experiment, Matthias Fassl from the research group of CISPA-Faculty Dr. Katharina Krombholz examined why users rarely take these extra steps. The results were published in a paper at the Conference on Human Factors in Computing Systems in April 2023.

On the difficulties of performing authentication ceremonies: A self-experiment

Matthias Fassl

For many years now, messenger services such as Signal, Threema, or WhatsApp have been among the most popular and widespread forms of digital exchange between people. Not only text messages can be exchanged, but also images, documents and voice messages – of both private and professional nature. This highlights the need for these services to be secure. Today, end-to-end encryption is the norm on many messenger services. This means that "as soon as messages leave a device, they are encrypted in such a way that only the receiving device can decrypt them", explains CISPA researcher Matthias Fassl. "The big uncertainty is whether the right person is actually sitting at the other end", Fassl continues. "One of the possible vulnerabilities is a man-in-the-middle attack, where someone pretends to be your friend Paul, for example. To fend off such an attack, users need to check that the key used to decrypt the text belongs to the right recipient. This is done with the help of authentication ceremonies." In practice, this means that two users meet and authenticate each other via QR codes displayed on their smartphones.

New methodological territory to cover a research gap

The challenge, however, is that users rarely perform authentication ceremonies. According to Fassl, one reason for this is that the concept underpinning end-to-end encryption is "trust-on-first-use". This concept assumes that users trust the contacts they add to their messenger and confirm this trust by contacting them via the messenger. The actual encryption of the chat then takes place in the background. For this reason, many people do not even know that only the actual authentication of their chat partners offers the greatest possible security. According to Fassl, there are hardly any figures or studies on how often users perform authentication ceremonies. This is precisely where Fassl's interest comes in: He wants to know what makes these ceremonies so difficult to carry out. "Over time, it occurred to me that maybe factors that don't affect the user interface but that relate to our interaction with each other can also play a role", he says.

For his study, Fassl chose the method of autoethnography. “Ethnographic approaches are relatively practical for studying social and cultural factors between multiple people interacting socially”, he says. “Autoethnography is the same thing, but with your own person. It’s a special case and not as popular because the person doing the study and the person being studied are one and the same.” Nonetheless, he says, there are also advantages to an autoethnographic approach since “you don’t always have to record everything with pinpoint accuracy, because things can be added afterwards from memory.” The choice of method, however, created a challenge when it came to publishing the results. “Due to the fact that the method is not so popular, it was a bit difficult to publish. It was also only the second autoethnography I found in the cybersecurity field.”

Underestimated effort to perform the ceremonies

All the more interesting – even in the eyes of the reviewers – were the results Fassl was able to compile over several months of self-observation. For example, he was able to prove that the biggest challenge to authentication ceremonies is the planning and organizational effort. “Not only do I have to meet with people, but I also have to figure out how to fit the ceremony into the conversation”, the Fassl explains. “Personally, for the study, I went through all my contacts in messengers and looked to see where there was already a green check mark and who I still needed to authenticate. I then tried to work through that relatively systematically.” In this process, even before the actual ceremony, he says there are several points where breaks can occur. “Those are moments when people drop out because they forget about the ceremony or more exciting topics of conversation come up during the social interaction.”

He also often had to explain to his dialogue partners what the authentication ceremony was all about. Here it became apparent that personal factors play a decisive role, which can vary greatly between individuals. “In my case, the strongest influence has been that my contacts know that I work in academia in the field of cybersecurity”, he says. “This means when I suggest trying a security mechanism, there’s a lot of authority that comes with it. Dissent toward me would express a lack of regard.” The importance of framing one’s own experience becomes apparent in an autoethnographic approach, as he continues: “What I described in the study was probably still a relatively positive outlook because of my personal factors. Other people probably would have had a much harder time performing these ceremonies.”

Consequences and possible design changes

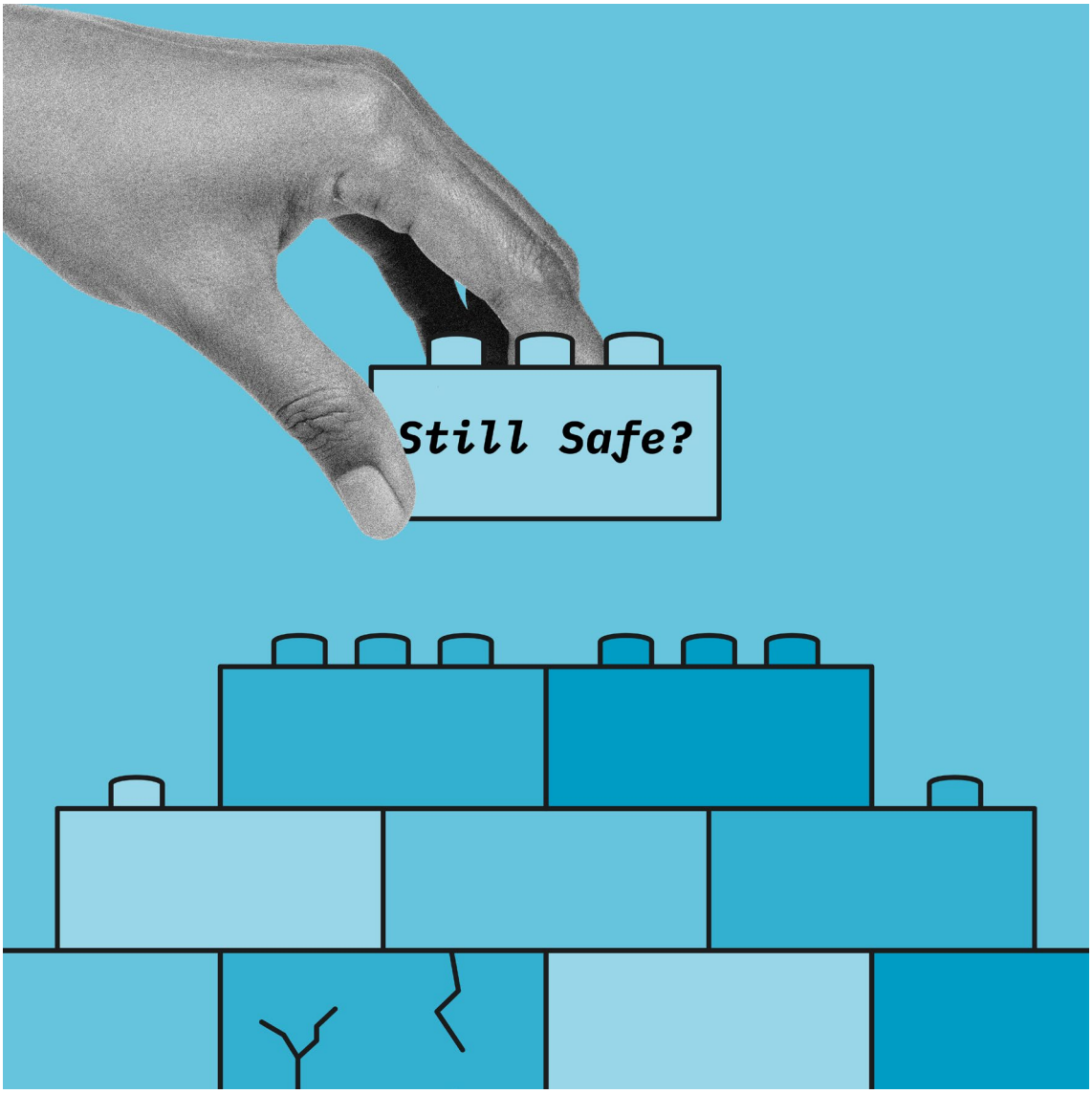
More fundamentally, Fassl is keen to emphasize the close intertwining of safety issues with human factors.

“I believe that behind every technical security there is actually always a human factor that wants to protect something or avoid something.” This is particularly important in authentication ceremonies, he adds. “The difference between normal security mechanisms and authentication ceremonies is that we often implement the former just for ourselves. The latter are a special case in that we need to work together to ensure our collective security.” Technology is only ever part of the solution, which is what the term ‘socio-technical gap’ attempts to describe. The socio-technical gap describes the difference between what the technology offers and what users are able and willing to implement. “In this case, it means that the authentication ceremony must somehow be incorporated into everyday life and conversations”, Fassl explains.

“My thoughts go in the direction of relieving the user of organizational effort”, he continues. “Technical support could help bridge the social-technical gap. This would be possible, among other things, with automated notifications on smartphones at appropriate times. Individuals could still opt out, of course. But it would be a convenient reminder.” Fassl and his colleagues have many ideas for trying out new solutions themselves. And even if no concrete research project is currently linked to it, Fassl is sure: “I don’t think the topic is dead yet.”

Fassl, Matthias; Krombholz, Katharina (2023) Why I Can’t Authenticate – Understanding the Low Adoption of Authentication Ceremonies with Auto-ethnography. In: CHI23, 23-28 April 2023, Hamburg, Germany. Conference: CHI International Conference on Human Factors in Computing Systems

Researcher: Matthias Fassl
Author: Felix Koltermann



© Lea Mosbach

Alexander Dax received two Distinguished Paper Awards at the prestigious USENIX Security Symposium 2023. Dax, who is a PhD student and CISPA researcher, is excited to receive so much encouragement from the research community. He received one of the two coveted awards for his paper "Hash gone bad: Automated discovery of protocol attacks that exploit hash function weaknesses". In this paper, Dax shows that automated security analyses of Internet protocols are often inaccurate because they are based on false assumptions – in this case, perfect hash functions. He explains what hash functions are, what they are used for, and how he intends to use his research to improve the automated analysis of protocols.

Reality check for automated analysis of protocols

Alexander Dax

Various protocols are used to ensure that data can be sent back and forth securely on the Internet. They regulate who can send what to whom, when, and in what form. One of the best-known Internet protocols in continuous use is TLS, short for Transport Layer Security. TLS primarily regulates how communication between web applications is encrypted. For example, browsers such as Google Chrome and Mozilla Firefox communicate with a web server every time a website is accessed. To prevent this communication from being infiltrated by attackers, a secure connection must first be established before the actual communication can take place. This step ensures that the communication partners are who they claim to be and that no third party can intervene. Once this has been clarified, cryptographic keys can be exchanged securely, thus enabling confidential communication. But how can this be done securely?

Hash functions as a security guarantee

"Almost every security protocol uses hash functions", Dax explains. They enable the creation of a check value and thus of a kind of digital fingerprint. This can be used to check whether data has been manipulated on its way from A to B. "These functions take any value, of any size, and turn it into a smaller value with a fixed size", he elaborates. That alone doesn't do much; the functions must also have certain properties. "These include the fact that specific data contents, such as passwords, must always result in the same value when calculated with the same hash function. Conversely, however, it must not be possible to infer the data content from the hash value." Another important property of hash functions is that different source data must not be converted to the same hash value. "We talk about collisions when that happens", Dax says. And this is where theory and practice get in each other's way: "In reality, there are no perfect hash functions. It's always just a matter of time before there are collisions. In addition, the state of technology has changed. With old hash functions, it is now possible, with enough computing power, to try out different values until the original value for a hash value is found. This is called a brute force attack."

Networks must be future-proof

Such attacks are very difficult to execute on modern hash functions, according to Dax, so they have not been

an everyday problem so far. “However, technology is evolving very quickly, and we need to make sure our networks are future-proof as well.” And that’s where Dax’s research around tools for automated protocol security analysis comes in. “It’s not enough to say that a protocol is secure. We also need to be able to prove it formally. That means we need precise mathematical definitions of how the protocol behaves, to enable us to then calculate how secure it is.” These testing procedures are enormously costly, which is why they have now been automated. “There are tools for this, such as Tamarin Prover or Proverif, that can do the work for us. The problem is: So far, these tools often only work with model representations of hash functions, which are perfect in this form. But in practice, we know that they often just aren’t.”

Too perfect isn’t any good either

Acknowledging these imperfections is the first step to improving the tools. It also has another benefit: “We have modeled different variants of weak hash functions and built them into Tamarin Prover and the Proverif tool. By doing this, we also want to find out how big the impact of different weaknesses in the hash functions is on the overall security of the protocol.” Formal security proofs of protocols are not some nerdy researcher’s stuff, but they have

»Technology is evolving very quickly, and we need to make sure our networks are future-proof as well.«

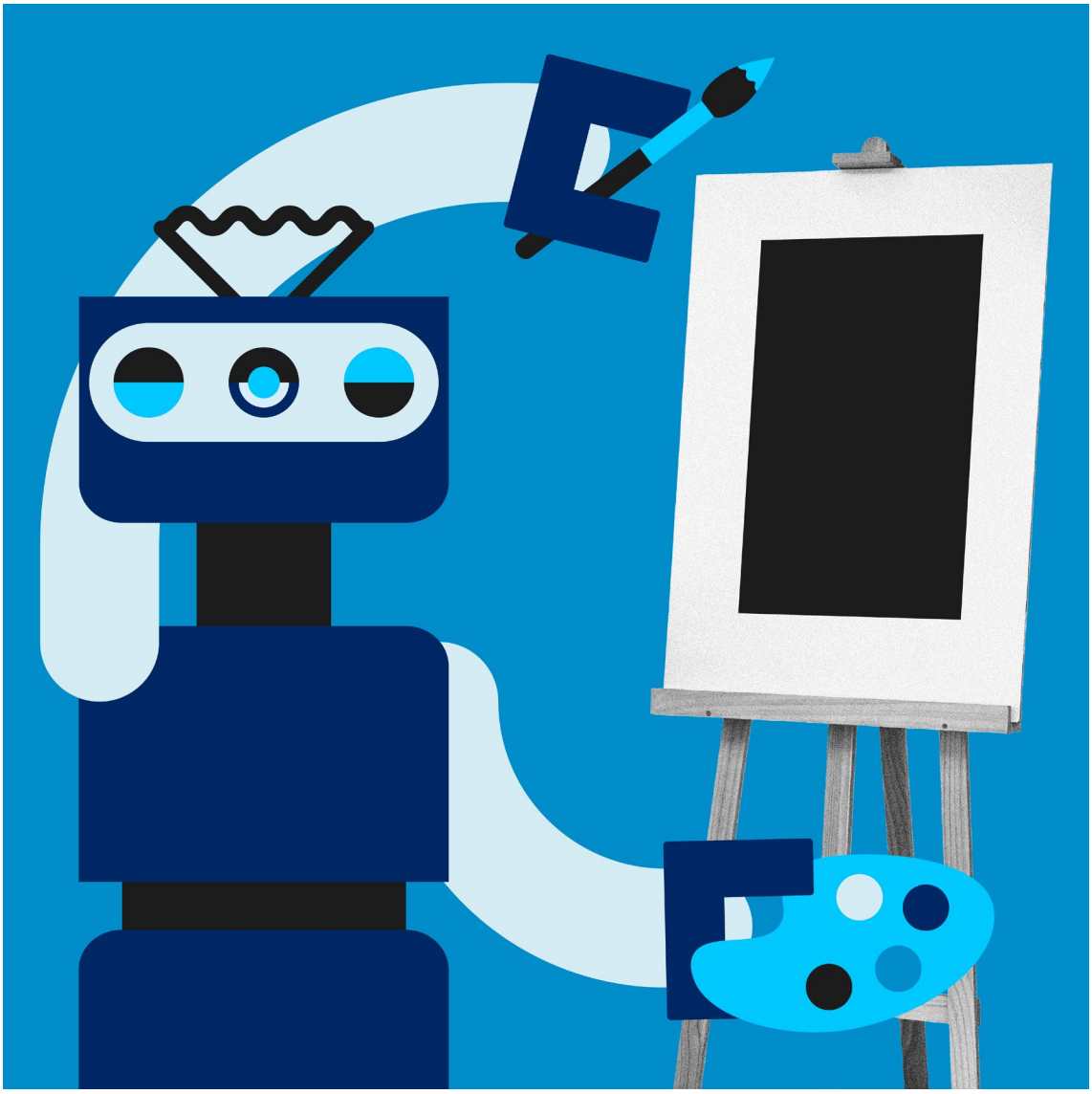
long since arrived at the world's major tech companies as well. "Many large companies, such as Google, employ cryptographers to check how secure the protocols they use are. This is very time-consuming to do manually, and even checking automated security analysis currently requires a lot of effort. We want to make the tools better so that in the future this will require significantly less manpower and effort, and automated checking can guarantee real protocol security."

At the source

Dax works with the group of CISPA-Faculty Professor Dr. Cas Cremers, and is thus at the source of research questions related to automated protocol auditing. Several years ago, Cremers and his colleagues developed the Tamarin Prover, which is used by companies such as Mozilla and Amazon. "My research is part of a larger project to improve automated security analysis. I've been working on it for years. That my research on hash functions has now resulted in an excellent paper is great," Dax says. He is now a CISPA veteran of sorts. "I've been involved since 2016. I started as a student assistant with Michael Backes, then in Robert Künnemann's group, and now I'm a PhD researcher working with Cas. Somehow, I've grown along with CISPA."

Cheval, Vincent; Cremers, Cas; Dax, Alexander; Hirschi, Lucca; Jacomme, Charlie; Kremer, Steve (2023) Hash Gone Bad: Automated discovery of protocol attacks that exploit hash function weaknesses. In: 32nd USENIX Security Symposium, 9-11 Aug 2023, Anaheim, CA, USA. Conference: USENIX Security Symposium

Researcher: Alexander Dax
Author: Annabelle Theobald



© Lea Mosbach

In the past year, AI image generators have experienced unprecedented popularity. With just a few clicks, all kinds of images can be generated: even dehumanizing imagery and hate memes. CISPA researcher Yiting Qu from the team of CISPA-Faculty Dr. Yang Zhang has analyzed the proportion of these images among the most popular AI image generators. She has also explored how their creation can be prevented with effective filters. She presented her paper "Unsafe Diffusion: On the Generation of Unsafe Images and Hateful Memes From Text-To-Image Models" at the ACM Conference on Computer and Communications Security in Copenhagen in November 2023.

Newly developed filter to prevent AI image generators from distributing 'unsafe images'

Yiting Qu

When people talk about AI image generators, they are often talking about so-called text-to-image models. This means that users can prompt the generation of a digital image by entering text prompts into an AI model. The textual input determines not only the content of the image, but also the style. The more extensive the training material of the AI image generator has been, the more possibilities of image generation ensue. Among the best-known text-to-image generators are Stable Diffusion, Latent Diffusion, and DALL·E. "People use these AI tools to draw all kinds of images", CISPA researcher Yiting Qu says. "However, I have found that some also use these tools to generate pornographic or disturbing images, for example. So the text-to-image models carry a risk." It becomes especially problematic when these images are shared on mainstream platforms, where they experience widespread circulation, she adds.

The notion of 'unsafe images'

Qu and her colleagues use the term 'unsafe images' to refer to the fact that AI image generators can be prompted to generate images of inhumane or pornographic content with a few simple instructions. "Currently, there is no universal definition in the research community of what is and is not an unsafe image. Therefore, we took a data-driven approach to define what unsafe images are", Qu explains. "For our analysis, we generated thousands of images using Stable Diffusion", she continues. "We then grouped these and classified them into different clusters based on their meanings. The top five clusters include images with sexually explicit, violent, disturbing, hateful and political content."

To concretely quantify the risk of AI image generators generating hateful imagery, Qu and her colleagues fed four of the best-known AI image generators, Stable Diffusion, Latent Diffusion, DALL·E 2, and DALL·E mini, with specific sets of hundreds of text inputs called prompts. These sets of text inputs came from two sources: the online platform 4chan, popular in far-right circles, and the Lexica website. "We chose these two because they have been used in previous work investigating unsafe online

content", explains Qu. The goal was to find out whether or not the image generators produced so-called 'unsafe images' from these prompts. Across all four generators, 14.56 percent of all generated images fell into the 'unsafe images' category. At 18.92 percent, the percentage was highest for Stable Diffusion.

Filter functions block image generation

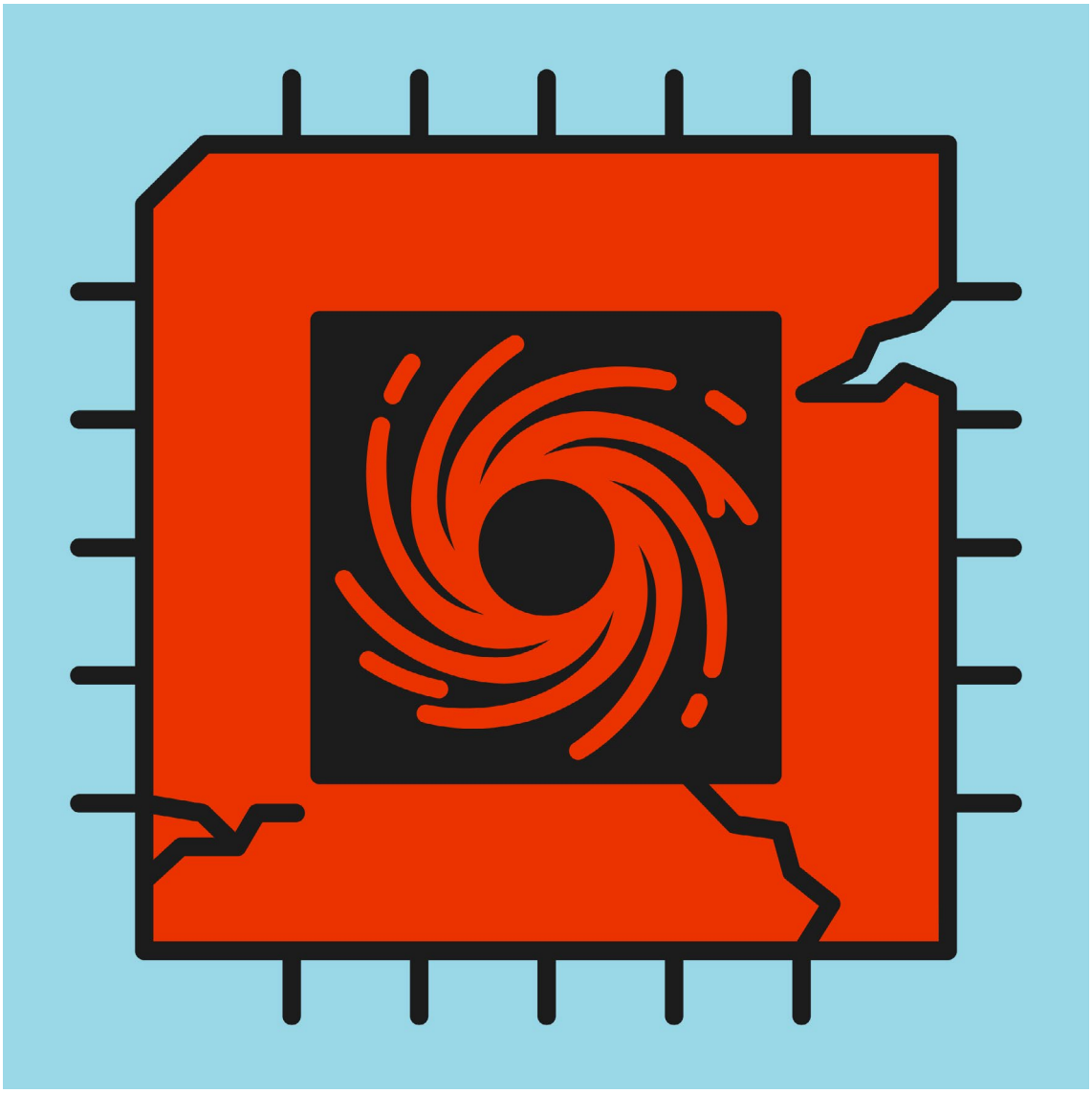
One way of preventing the spread of inhumane imagery is to program AI image generators to not generate this kind of imagery in the first place or to not output it. "I can use the example of Stable Diffusion to explain how this works", Qu says. "You define several unsafe words, such as nudity. Then, when an image is generated, the distance between the image and the word defined as unsafe, such as nudity, is calculated. If that distance is less than a threshold, the image is replaced with a black color field." The fact that so many unsafe images were generated in Qu's study of Stable Diffusion shows that existing filters do not do their job adequately. The researcher therefore developed her own filter, which scores a much higher hit rate in comparison.

However, preventing image generation is not the only option, as Qu explains. "We propose three remedies that follow along the supply chain of text-to-image models. First, developers should curate the training data in the training or tuning phase, i.e., reduce the number of unsafe images." This is because 'unsafe images' in the training data are the main reason why the model poses risks later on, she says. "The second measure for model developers is to regulate user-input prompts, such as removing unsafe keywords." The third possibility concerns dissemination after image generation, Qu adds: "If unsafe images are already generated, there must be a way to classify these images and delete them online." For the latter, in turn, there would need to be filtering functions for the platforms on which these images circulate. With all these measures, the challenge is to find the right balance. "There needs to be a trade-off between freedom and security of content. But when it comes to preventing these images from experiencing wide circulation on mainstream platforms, I think strict regulation makes sense", Qu says. She hopes to use her research to help reduce the number of harmful images circulating on the Internet in the future.

»If unsafe images are already generated, there must be a way to classify these images and delete them online.«

Qu, Yiting; Shen, Xinyue; He, Xinlei; Backes, Michael; Zannettou, Savvas; Zhang, Yang (2023) Unsafe Diffusion: On the Generation of Unsafe Images and Hateful Memes From Text-To-Image Models. In: CCS 2023, 26-30 Nov 2023, Copenhagen, Denmark. Conference: CCS ACM Conference on Computer and Communications Security

Researcher: Yiting Qu
Author: Felix Koltermann



© Lea Mosbach

AMD developed so-called Secure Encrypted Virtualization (SEV) with the primary goal of making its cloud services more secure. But until recently, even the latest versions of the security feature, SEV-ES (Encrypted State) and SEV-SNP (Secure Nested Paging), were vulnerable to a software-based attack. This was discovered by CISPA researcher Ruiyi Zhang, who is a member of the research group of CISPA-Faculty Dr. Michael Schwarz. He constructed an attack called CacheWarp, which in the worst case gives attackers comprehensive access rights to data and even the possibility to manipulate it - something that previous attack techniques did not achieve. Zhang received support from his research team at CISPA and Andreas Kogler from Graz University of Technology. According to AMD, the vulnerability has now been fixed with an update.

Vulnerability in AMD security feature detected

Ruiyi Zhang

Large cloud platforms are booming. “Cloud services offer companies the flexibility to purchase computing power and storage space whenever they need it”, explains Ruiyi Zhang. The security of these services is crucial but has been compromised in the past by the discovery of vulnerabilities and the potential for attacks. “Cloud services are based on so-called virtualization, which can save hardware components and, as a result, staff”, he says. According to Zhang, virtualization means the creation of several virtual machines on a single physical computer. A virtual machine is basically a software-based computer equipped with everything a regular computer has: its own memory, a CPU, and an operating system. Virtualization can thus make many computers out of one single computer, provided it has the necessary computing power.

Security feature with vulnerabilities

The distribution of resources and the corresponding separation of processes is handled by the hypervisor. This software distributes resources such as memory and computing power and isolates the operating systems. The hypervisor thus acts as a kind of host for the virtual machines. To prevent it from becoming a point of attack, the processor manufacturer AMD introduced the first generation of Secure Encrypted Virtualization (SEV). The idea behind SEV: The memory of each virtual machine is encrypted with a separate key, which is supposed to prevent overlapping data access and access by an untrustworthy hypervisor or one that has been taken over by attackers. “Several security vulnerabilities were quickly identified. In addition, SEV and SEV-ES initially used encryption without an identity check, which allowed for data manipulation. Also, not all parts of the memory were encrypted”, explains CISPA-Faculty Michael Schwarz. Schwarz is an expert for security vulnerabilities in CPUs and was involved in the discovery of, for example, Spectre, Meltdown and ZombieLoad. AMD reacted to the problems by further developing SEV into the features SEV-ES (Encrypted State) and, most recently, SEV-SNP (Secure Nested Paging). According to AMD, SEV-SNP provides a strong memory integrity, which should prevent hypervisor attacks.

A few lines of code

About half a minute, access to a server room and a few lines of code is all that Zhang would need to gain access to every virtual machine and to view and modify

everything he wants with administrator rights. Finding out how exactly this is possible took several months of work. “To the best of our knowledge, CacheWarp is the only software-based attack so far that can defeat SEV-SNP like that”, Zhang explains.

»To the best of our knowledge, CacheWarp is the only software-based attack so far that can defeat SEV-SNP like that.«

A computer travels through time

“First, we need to be able to log into a system. For this purpose, we employ a method that we called TimeWarp”, Schwarz says. This method exploits the fact that in certain scenarios, computers memorize which code they need to execute next. “We can reset what the computer has memorized as the next step. This makes the computer execute code that it executed before because it reads an outdated so-called return address from memory. The computer thus travels back in time. However, the old code is executed with new data, which leads to unexpected effects. If you use this method cleverly, you can change the program logic”, Schwarz explains. Zhang adds: “TimeWarp thus allows us to change the program logic in a virtual machine such that we can log in without knowing the password.”

0 represents success

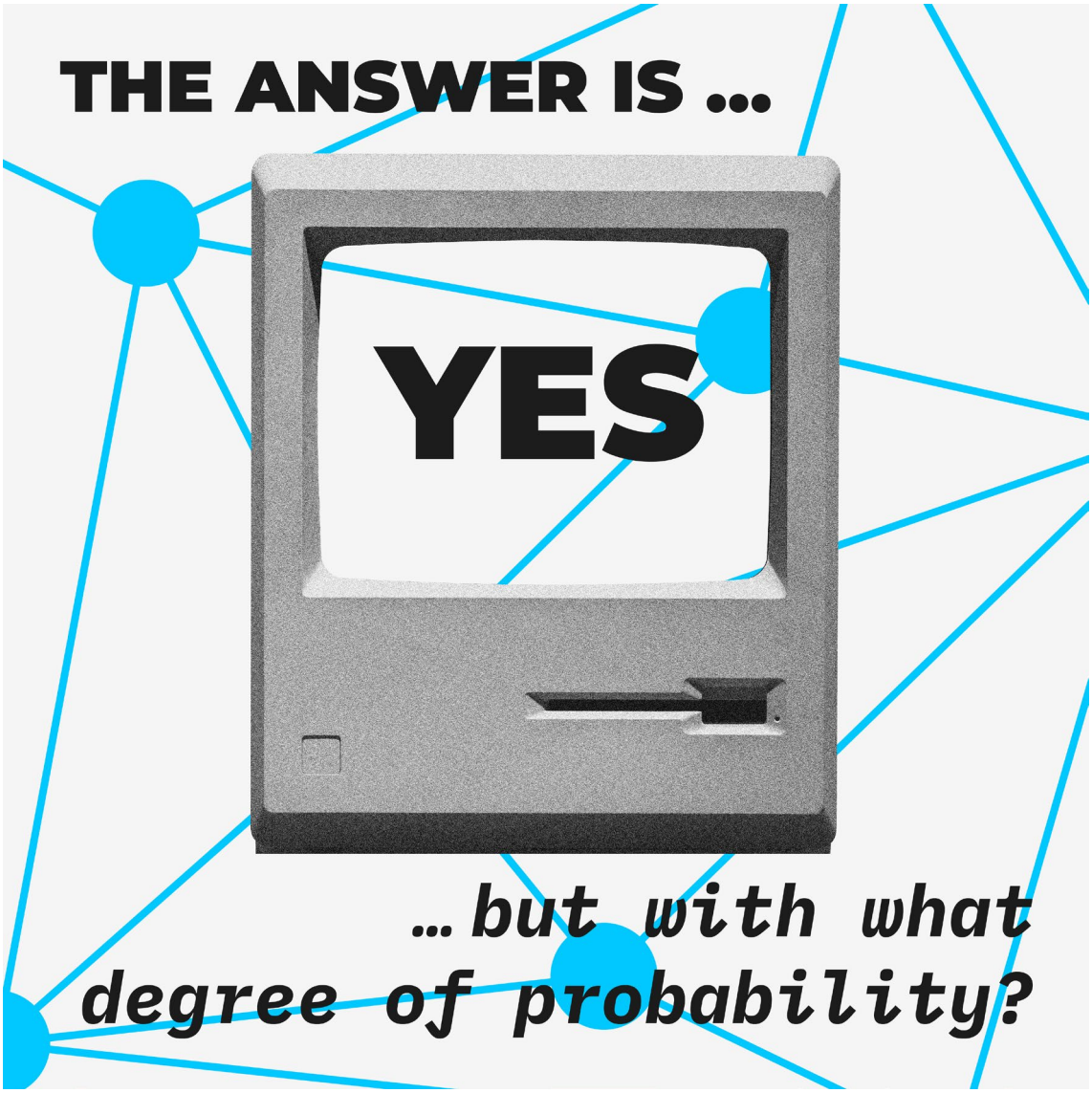
Combined with another method called Dropforge, it is also possible to manipulate the cache and reset changes made on data. “Even if it doesn’t seem intuitive, this even allows you to be granted administrator rights. This is achieved by exploiting details of the program logic”, Schwarz says. In computer science, ‘0’ often represents success, whereas other values represent potential error codes. According to Schwarz, ‘0’ is also the default value for data if no different value is stored. “When the system

tests whether the respective user is an administrator or not, the query will return ‘0’ if you are an administrator. If you are not an administrator, a different value will be returned. With Dropforge, this return value can be reset. No matter if you are an administrator or not, the memory will contain the initial value of ‘0’. The system then assumes that you are an administrator”, Schwarz explains. “With this combination, we have unlimited access to the virtual machine”, Zhang adds.

Trust is good

In their paper “CacheWarp: Software-based Fault Injection Selective State Reset”, Zhang and his fellow researchers not only describe the attack methods but also suggest a compiler-based solution to mitigate the attacks. In addition, they want to provide an open source testing tool for the vulnerability. “We don’t want to rely on the statement that something is secure. We want to be able to verify it”, Schwarz explains. Since discovering CacheWarp, the researchers have been in communication with AMD: The manufacturer has indicated to them that the vulnerability has been fixed by now.

Zhang, Ruiyi; Gerlach, Lukas; Weber, Daniel; Hetterich, Lorenz; Lü, Youheng; Kogler, Andreas; Schwarz, Michael (2024) CacheWarp: Software-based Fault Injection using Selective State Reset. In: 32nd USENIX Security Symposium, 9-11 Aug 2024, Anaheim, CA, USA. Conference: USENIX Security Symposium



© Lea Mosbach

For a machine learning algorithm to be trustworthy, the end user needs to know how confident the model is about each of its predictions. To date, accuracy has been a major criterion for the evaluation of a model. However, this does not reflect on how confident the model is in processing each input. Scientists have devised methods to quantify the 'uncertainty' but most of these methods are computationally expensive. The task becomes even more challenging when inputs are given in the form of networks with meaningful connections such as happens, for example, in drug discovery, medical diagnosis, or traffic forecast. In their paper "Conformal Prediction Sets for Graph Neural Networks", CISPA researcher Soroush H. Zargarbashi and his colleagues have successfully tested a new method to define a set of possible predictions for these networks that is guaranteed to include the true prediction.

New method for uncertainty quantification in machine learning applications

Soroush Zargarbashi

Many machine learning applications are based on graph neural networks (GNNs). “In many real-life scenarios, we deal with graphs. There are meaningful connections between our datapoints, and with GNN we take those connections into account”, CISPA researcher Soroush H. Zargarbashi explains. Graphs are a type of abstract data structure consisting of two elements, nodes and connections between nodes, called edges. Graphs can, for example, provide models for social networks, sensor networks, scientific papers with their references, etc. There is, however, one particular characteristic that causes problems in certain areas of application, Zargarbashi continues: “If you use a model as a black box, an input always results in an output like your car seeing the scene and deciding to steer left. But if you don’t know if the model is sure about this particular output, it becomes highly untrustworthy especially in safety-critical domains where the user needs an uncertainty estimate of the model.” The problem is that the models often overestimate their prediction quality, while underestimating the uncertainty factor of the predictions.

»If you use a model as a black box, an input always results in an output like your car seeing the scene and deciding to steer left.«

To illustrate how important it is for models to provide reliable uncertainty estimates for their predictions, Zargarbashi gives an example: “Imagine you are using a medical diagnostic system to decide whether a patient has a certain disease. In this case, it is very important that your model can predict this with a high degree of certainty. If the model cannot do this, further diagnoses have to be made. The idea behind the quantification of uncertainty is to refine these predictions.” Consider, for example, procedures in which AI is used to determine whether an organ is cancerous or not by analyzing MRI images automatically. Here, it is important to know the prediction quality for each individual input. In other words, an accuracy of 90 percent can be very risky for the remaining 10 percent.

Conformal prediction as a solution for uncertainty quantification

“There are methods to quantify this uncertainty. However, they are computationally expensive, hard to apply and worst of all, many of them don’t work on graphs”, Zargarbashi says. Many of these methods require modifications to the model architecture or retraining of the model. “However, there is growing interest in an alternative approach known as conformal prediction”, he continues. Conformal prediction (CP) is a statistical method for creating prediction sets that has been known since the late 1990s and that does not require any assumptions about the prediction algorithm. Zargarbashi mentions that CP can work around the model like a kind of wrapper and generate a set of possible predictions with a user-defined probability guarantee for the correct answer. But how exactly does it work? “For example, for a new patient, you tune the algorithm to produce sets that guarantee the true answer with a probability of 95 percent. This works for any model, even those who are 60 percent accurate. You just need a random sample of previous patients with their correct diagnosis. In this way, for each patient, we have a set of possible diagnoses which we know includes the correct answer with very high probability.”

Making machine learning trustworthy

Basically, the method developed by Zargarbashi and his colleagues is an alternative to uncertainty quantification that not only works on graphs (data with connections) but also uses the information from the relations between datapoints. Their method is computationally inexpensive and easy to implement, provided that additional data are available. The key advantage of this approach, according to Zargarbashi, is that CP is “model independent, meaning that it doesn’t matter which model is used, so you don’t need to train anything from scratch.” To further improve the applicability of this study, they develop a method called Diffusion Adaptive Prediction Sets, which uses the connections between datapoints to

improve the uncertainty estimation quality. In the published paper, the detailed empirical analysis of this method is embedded in a comprehensive theoretical investigation of when CP can be applied to GNNs. With their study, Zargarbashi and his colleagues make an important contribution to increasing the trustworthiness of machine learning models on graph data.

»There are methods to quantify this uncertainty. However, they are computationally expensive, hard to apply and worst of all, many of them don't work on graphs.«

Zargarbashi, Soroush H.; Antonelli, Simone; Bojchevski, Aleksandar (2023) Conformal Prediction Sets for Graph Neural Networks. In: Proceedings of the 40th International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202. Conference: International Conference on Machine Learning

ABOUT CISPA

The CISPA Helmholtz Center for Information Security is a national Big Science institution within the Helmholtz Association of German Research Centers. CISPA researchers explore all aspects of information security. They conduct cutting-edge foundational research as well as application-oriented research, addressing the most pressing challenges in cybersecurity, artifical intelligence and privacy. Research results achieved at CISPA find their way into industrial applications and products that are available worldwide. CISPA thus contributes to German as well as European competitiveness.

CISPA offers a world-class research environment as well as extensive resources to a large number of researchers. It strongly supports the undergraduate and graduate education of cybersecurity students and seeks to become an elite training ground for the next generation of cybersecurity experts and leading scientists in this field. CISPA is located in Saarbrücken and St. Ingbert. The center's proximity to France and Luxembourg puts it in an ideal position for cross-border cooperation with other research institutions.

Our research currently focuses on the following six research areas:

Algorithmic Foundations and Cryptography

Trustworthy Information Processing

Reliable Security Guarantees

Threat Detection and Defenses

Secure Connected and Mobile Systems

Empirical and Behavioral Security

IMPRINT

CISPA – Helmholtz-Zentrum
für Informationssicherheit gGmbH
Stuhlsatzenhaus 5
66123 Saarbrücken, Germany

Publisher

Sebastian Klöckner

Editor-in-Chief

Felix Koltermann,
Eva Michely,
Annabelle Theobald

Editors

Lea Mosbach,
Janine Wichmann-Paulus

Illustration

Janine Wichmann-Paulus

Design

Stephanie Bremerich,
Tobias Ebelshäuser

Photography

May 2024

Information as of

T: +49 681 87083 2867
M: pr@cispa.de
W: https://cispa.de/

Contact
Corporate
Communications

New approach improves automatic detection of vulnerabilities in processors

Why visual digital certificates are only secure in theory (so far)

Developing an open-source prototype for 2-factor authentication

New specification language is a game changer for automated software testing

A new a token-based system for humanitarian aid distribution combines accountability and privacy

The new gold standard: Rethinking Differential Privacy

Key management is a challenge for crypto funds

Website operators take security more seriously than data protection

Space oddities: Examining satellite security

Collide+Power:New side-channel attack affects all CPUs

MobileAtlas: Mapping mobile communications security

A new standard? Using web archives for live analyses of website security

Testing a new technique to safeguard against deepfakes

On the difficulties of performing authentication ceremonies: A self-experiment

Reality check for automated analysis of protocols

Newly developed filter to prevent AI image generators from distributing “unsafe images”

Vulnerability in AMD security feature detected

New method for uncertainty quantification in machine learning applications

CISPA
HELMHOLTZ CENTER FOR INFORMATION SECURITY